WHAT IS SIMPLE IS USABLE: IS BEAUTY TRULY A PREDICTOR OF PERCIEVED USABILITY?

A Thesis

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Master of Science

by

William George Miner

August 2011

ABSRACT

For more than a decade, the correlation between beauty and perceived usability has been a popular topic of focus in both Human Computer Interaction (HCI) and Human Factors (HF) circles. As opposed to previous research that has mainly looked at how attractive designs can increase usability (Tractinsky, 2000), this thesis explores the directionality of the correlation between beauty and perceived usability. Specifically, to what extent does attractiveness increase and unattractiveness decrease perceived usability?

It was hypothesized that a bad design would have more of an impact on perceived usability than a good one. Thus, based on the phenomenon of "negativity bias" (Baumeister et al., 2001; Rozin & Royzman, 2001; Griffin & Langlois, 2006), an experiment was designed to test if an "ugly" design had more of an effect on usability, in the negative direction, than a "beautiful" design had in the positive direction. Additionally, it was hypothesized that this perception of usability may change over time. Similar to the methods used by Tractinky (2000) and Soderegger & Sauer (2010), measures were taken before and after using a blogging website.

A preliminary study (N=48) was conducted to determine which of the 12 researcher generated layouts were the most beautiful, ugly, and aesthetically neutral. The 3 layouts that were selected – "beautiful", "ugly", and "neutral" – were then applied to three different blogs that had identical structure, content, and usability.

In a second, separate "main study" experiment, participants (N=69) were first presented with a static image of their assigned blog/beauty level. They then completed an abbreviated version of the System Usability Scale (SUS) and the Standardized Universal Percentile Rank Questionnaire (SUPR-Q). Next, the users were allowed five minutes to "surf" the website before

they completed a second SUS and SUPR-Q. Overall, the main study experiment was a between subjects 3 x 2 factorial design with the independent variable of aesthetic level (beautiful, ugly, neutral) crossed with the independent variable of time (pre-use and post-use). The dependent variable was usability, as recorded by a System Usability Scale (SUS) and Standardized Universal Percentile Rank Questionnaire (SUPR-Q).

The results of the experiment did not support Tractinsky's (2000) findings that the more beautiful a design is, the higher it's perceived usability will be. Rather, the "neutral" blog received significantly higher ($p<.05$) ratings of perceived usability that both the beautiful and ugly designs – which had significantly similar ($p>.05$) rated usability. In agreement with Tractinsky (2000) however, there was a significant increase in rating between pre and post use of the blogs.

Overall, these results suggest that beauty may not intrinsically influence perceived usability. In fact, it may be that simplicity – as seen in design of the "neutral" blog – is the true influencer of perceived usability and "beautiful" designs are merely predisposed to be more "simple" than ugly ones. Because of these and other novel results, further research is needed to tease out the true effect of beauty and simplicity on usability.

BIOGRAPHICAL SKETCH

William George Miner was born and raised in Silver Spring, Maryland and graduated from Gonzaga College High School in May 2006. As a student at Cornell University, he majored in Design and Environmental Analysis with a concentration in Human Factors and Ergonomics. During his time as an undergrad, he was a member of the Varsity Lightweight Rowing team and Psi Upsilon Fraternity brother. After earning his B.S. in May 2010, he continued his education at Cornell by pursuing a M.S. in Human Environment Relations with a concentration in Human Factors and Ergonomics. As a graduate student, he earned a minor in Information Science and was selected to be a State University of New York Diversity Fellow.

Dedicated to my parents – Bill and Allison

ACKNOWLEDGEMENTS

In the writing of this thesis, there are a number of people who I am in indebted to for their time, effort, and encouragement.

First, I would like to thank Cornell University and the State of New York for their generous financial contribution to my education. For a stressed and sleepless college student, being grated monetary piece of mind is an enormous aid. Additionally, I would like to thank the Design and Environmental Analysis Department for its monetary contributions to my research.

Next, I would like to express gratitude to all of my educators, past and present. Without your guidance and insight, this thesis would have been inconceivable. I would especially like to acknowledge, Professors Jack Elliot, Jan Jennings, Alan Hedge, and Gary Evans, whose instruction has been especially influential.

I would also like to sincerely thank all of my friends and family for your love, support, and most of all humor. Your encouragement has always been one of my strongest driving forces and this thesis is a testament to your own personal qualities. Of special mention are my grandparents – Millie, Owen, "Mootie" & "Pootie" – and my sister Ava.

I would like to extend my deepest gratitude to my minor member Professor Dan Cosley. Without, your introductory HCI class, it is a certainty that not just this thesis, but my entire research and professional direction would have been different and assuredly less successful.

Finally, I would like to acknowledge the enormous time and effort devoted to this thesis by my committee chair Professor David Feathers. David's enormous knowledgebase, helpful

suggestions, and positive demeanor have been an integral part of this thesis's success. For his mentorship, I am exceedingly grateful.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Informing Design Through Pleasure

In the brief history of the World Wide Web, there have been several, well-documented "showdowns" between websites that offer similar content. For instance, social news sites, Digg.com and Reddit.com, compete for the same users, and seek to distinguish themselves through a number of factors (TechCrunch, 2010). Most notably with their designs, Digg uses a sleek, modern interface with bright colors, rounded corners, and a highly organized structure. Reddit on the other hand has a plain white background with dark blue and grey text. If a purely functional evaluation of the differences between these websites is used, you might conclude that one simply had better content verses the alternative. However, a growing body of literature suggests that usability may be one the factors that subconsciously create a preference for one over another.

Usability is defined as a quality of a man-made object that reflects both its ease of use and learnability (Lewis & Sauro 2009; Nielsen, 1993). A field dedicated to the study of this concept, Human Factors Engineering (HFE), analyses how these aspects inform the design of products. Over the last 20 years, however, pleasure, or more precisely positive affect, has been increasingly recognized as an important addition to the elements ease of use and learnability (Hancock, 2005).

P. A. Hancock (2005) coined the term "Hedonomics" as a way to describe this emerging branch of Human Factors and Ergonomics. Where traditional HFE is concerned with the

prevention of pain, Hedonomics is concerned with the promotion of pleasure. Related to Kansei Engineering – a method for engineering appropriate and beneficial emotions into products (Dai et al., 2011) – these two theories inform design because the:

> "Feelings and impressions of a product are important for the decision of purchasing it or not. Designing attractive products therefore requires knowledge about the feelings and impressions the products evoke on the customer and the user." (Schütte, 2005, p. i)

The world of advertising has long known of this effect. Television commercials, for instance, often try to evoke certain feelings in the viewer that will in turn influence behavior. Research has empirically supported this claim and emphasizes how strongly subconscious feelings affect our beliefs about a product (Edell & Burke, 1987).

Overall, what these theories of emotional and affective design espouse is the need to approach engineering with a more humanistic approach. Too often, products are created that are sterile and emotionless. Hedonomics and Kansei Engineering recognize that while the prevention of the negative is important, the promotion of the positive can elevate a design from ordinary to extraordinary. Norman, in his book Emotional Design (2004), hypothesizes that "attractive things make people feel good, which in turn makes them think more creatively" which in turn makes "it easier for people to find solutions to the problems they encounter" (p. 19). By paying attention to human "pleasure", engineered objects can promote well-being, efficiency and effectiveness. As noted by Norman, a conspicuous way in which "pleasure" is generated is through attractiveness; thus, this thesis will look at how it directly affects usability.

In the development of all products, important decisions have to be made about what portion of resources should be given to various aspects of the project. Often, one of the most

controversial is the decision of whether to hire a creative professional like a graphic or industrial designer. By looking at how the aesthetics of a product affects user's overall satisfaction, this research may begin to expose the effects this type of expertise can have during a product's development.

1.2 The Role of Beauty

In 1790, on the subject of beauty, philosopher Immanuel Kant wrote:

> As regards to the agreeable, everyone acknowledges that his judgment, which he bases on a private feeling and by which he says that he likes some object, is by the same token confined to his own person. Hence if he says that canary wine is agreeable he is quite content if someone else corrects his terms and reminds him to say instead: It is agreeable to me...It is quite different (exactly the other way round) with the beautiful. It would be ridiculous if someone who prided himself on his taste tried to justify [it] by saying: This object (the building we are looking at, the garment that man is wearing, the concert we are listening to, the poem put up to be judged) is beautiful for me…if he proclaims something to be beautiful, then he requires that same liking from others; he then judges not just for himself but for everyone, and speaks of beauty as if it were a property of things. (Kant & Pluhar, 1987, p. 55)

In this passage, Kant addresses the differentiation between taste and beauty. He argues that taste is a subjective quality, existing solely within an individual's mind. Each person has their own tastes, likes and dislikes. Kant also argues that beauty is inherently objective and exists as a "property" of an object.

Our scientific understanding of beauty and aesthetics has come a long way since the Age of Enlightenment. However, much of what Kant postulated in the late 1700's still holds under our modern understanding of cognitive science. This section explores the psychological basis for

the human perception of beauty and aesthetics; describes how cognitive processes influence subconscious behavior; illustrates how beauty and aesthetics guided the development of fundamental aspects of society such as art and business; and relates the theory of beauty beyond human physical attractiveness and into the realm of aesthetics and Neuroesthetics. Finally, through an understanding of these concepts, beauty's most recent application, within the field Human Computer Interaction (HCI), will be studied.

1.2.1 Evolutionary Beauty and its Effects

In the Merriam-Webster Dictionary (2011), beauty is described as "the quality or aggregate of qualities in a person or thing that gives pleasure to the senses or pleasurably exalts the mind or spirit". While this is a somewhat nebulous definition, it has been a significant focus of psychology research to elucidate this "quality."

Starting with the evolutionary basis of beauty, it is important that we look at it in the context of human attractiveness. This core element of beauty is most readily explained by its connection to mate selection. As a central element of evolutionary theory, the concept of mate choice states that organisms will most often seek to reproduce with the most "genetically fit" and therefore fecund individual (Rhodes, 2006). By being attracted to and subsequently selecting the most fit individual, the organism in question (often the female), ensures that its offspring will in turn be "fit" and biologically successful within the environment (Rhodes, 2006). For example, in the animal world, male peacocks whose plumage exhibits that brightest coloration typically have the highest mating success rates (Loyau et al., 2007). This phenomenon of beauty as the basis for

genetic selection manifests itself in a number of ways in humans. While decidedly more complex than feather coloration, much of the same evolutionary reasoning is at work.

First is the issue of sexual dimorphism. Human females tend to prefer mates that exhibit stereotypical male traits (e.g. strong jaw, broad shoulders, etc.). In males, the same bias is true in that stereotypical female traits are preferred. This preference for certain traits, some seemingly arbitrary, is mostly linked to the organism's overall "fitness". For instance, taller and/or stronger males are assumed to be physically superior at stereotypical male tasks (hunting, protecting, etc.). Females who have typical youthful feminine characteristics (long hair or high waist-to-hip ratio) exhibit that they are more fertile and thus able to successfully bear and raise children. Other factors such as cultural and social success add another layer of complexity to the human mate choice phenomenon but transcend our focus of cognitive psychology.

Another important and overarching characteristic of physical attractiveness is the concept of Koinophilia (Edelman, 2008). This theory suggests that when choosing a mate, organisms prefer those that are a regression to the mean. That is to say, the closer one is to the species' "average" the more attractive they will be perceived (Rhodes, 2006). This phenomenon is exhibited especially well in the realm of facial attractiveness. First studied by Francis Galton in the late 1800s (Rhodes & Zebrowitz, 2002), averaging two faces together produces a third that is judged to be more attractive. Rhodes (2006), suggests that this theory of averageness may have an evolutionary basis, in that average features may have a functional advantage that suggests developmental stability.

Hand in hand with Koinophilia and averageness as beautiful traits, is the principle of symmetry. More so than averageness however, symmetry has a direct connection to physical and

genetic health. In humans, asymmetry often occurs as a result of "inbreeding, premature birth, psychosis, and mental defects" (Rhodes, 2006, p. 205). According to Rhodes (2006), as a signal of mating potential, symmetry rates amongst the strongest. Through these aforementioned traits of human beauty, we can begin to establish its biological and, in turn, evolutionary basis. These traits do not just influence mate selection but can also influence our decision making and general perception of others.

1.2.2 "What is Beautiful is Good"

Humans (and other animals) use physical appearance to determine the suitability of a potential mate. However, do humans use beauty to make other determinations about people? Dion & Berscheid (1972) answered "yes" to this phenomenon. In their influential paper, "What is Good is Beautiful", they proposed that in addition to perceiving beautiful people as superior mates, they also perceive them as having more positive attributes. In the paper, the authors write:

> The results suggest that a physical attractiveness stereotype exists and that its content is perfectly compatible with the 'What is beautiful is good' thesis. Not only are physically attractive persons assumed to possess more socially desirable personalities than those of lesser attractiveness, but it is presumed that their lives will be happier and more successful. (Dion & Berscheid, 1972, p. 289)

Overall, this conclusion extends the assumptions made under Koinophilia and symmetry, and asserts that physical attractiveness, on a fundamental level, affects how we inherently perceive others.

While this theory casts humans as a vain species, these findings have been confirmed in the real world. In a landmark study by Hamermesh & Biddle (1993), it was shown that earnings positively correlate to physical attractiveness. These results, which held across gender, found that ”unattractive” people earned significantly less than “average” looking people and “above average” looking people earned significantly more than those who were classified as “average” looking. Relating this back to the initial concept of positive affect, we see that beauty has a strong effect on things beyond mate selection. Of special note here however, is the finding that there was not so much a premium for being beautiful, as there was a penalty for being unattractive. Thus, there emerges the possibility of an additional “negative affect”.

Building on this finding, several researchers (Baumeister et al., 2001; Rozin & Royzman, 2001; Griffin & Langlois, 2006) have studied the phenomenon of people paying more attention to negative things than to positive ones. This theory, called “negativity bias”, was explored in depth by Griffin & Langlois (2006) in the context of human facial beauty. The authors found that, in agreement with Hamermesh & Biddle (1993), “unattractiveness [in humans is perceived as] ‘bad’ more than beauty is [perceived as] ‘good’” (p. 201). In all, these findings relate to Koinophilia and the “averageness” theory in that humans are more predisposed to notice defects in facial beauty, rather than appreciate exceptional symmetry or “averageness”. The evolutionary basis for this may be that, by noticing defects, first and foremost, individuals can more readily avoid those who are unsuitable for mating. Overall, the concepts of “Beautiful is Good”, “Bad is Stronger than Good” (Baumeister et al., 2001), and “Negativity Bias”, illustrate that the perception of beauty plays a central role in determining the success of individuals.

1.2.3 Neuroesthetics and its Correlates

It is important to note, that prior to now, we have only concerned ourselves with the concept of beauty – specifically human beauty. Also important, is the concept of aesthetics. Defined as "a particular taste for or approach to what is pleasing to the senses and especially sight" (Merriam-Webster, 2011), aesthetics extends past the concept of beauty as an evolutionary construct and attempts to answer the question "what makes art beautiful?"

1.2.3.1 "Artistic Creativity and the Brain"

Within the context of neurology and cognitive psychology, aesthetics presents an interesting parallel to human beauty. The emerging field of Neuroesthetics looks precisely at this parallel and aspires to ascribe a neurological foundation to the concept of aesthetics. Hallmark research by Semir Zeki provided much of the impetus for this field. In his seminal thesis on the topic, Zeki frames all art as a physical representation of our variable and visual brain (Zeki, 2001). The brain, by its very structure and purpose, is a constantly changing and evolving organ. In order to continuously adjust to the multitude of inputs brought to it by the rest of the body, the brain must be in a constant state of variation. In order to cope with these changes, Zeki writes that "art renders the destructive, isolating, and individualizing effects of variability safe in its pages, canvasses, and scores" (p. 51). Both the creation and perception of art, in a philosophical sense, allow the brain to both observe and experience itself. Further, "visual art contributes to our understanding of the visual brain because it explores and reveals the brain's perceptual capabilities" (p. 51).

With this somewhat inchoate explanation of the neurological origins of art, Zeki (2001), outlines two "supreme laws", or better stated, "supreme attributes", to guide the understanding of our "visual brain". First described is the "law of constancy". The author explains that our perception of the world is continually changing. Constancy allows the brain to categorize and understand what would be an otherwise indecipherable assault on our perceptual processing centers. Art, as Zeki (2001) hypothesizes, reflects this constancy by "registering the constant and essential characteristics of objects" (p. 52). The Mona Lisa, for example, is not an exact representation of a woman. It does however, distill the "essential characteristics" of one. Her subtle facial, features, flowing hair, and feminine figure all help to expose her female identity. By choosing these aspects to highlight, Da Vinci has locked into the painting as secondary level of emotional information.

Second, Zeki describes the "law of abstraction". Fundamental to how the brain works, the ability to abstract allows one to extend experience to new and novel stimuli. The author writes that:

> The capacity to abstract is also probably imposed on the brain by the limitations of its memory system, because it does away with the need to recall every detail. Art, too, abstracts and thus externalizes the inner workings of the brain. Its primordial function is thus a reflection of the function of the brain (Zeki, 2001, p. 52).

This means that abstraction – through all varieties of art – allows the mind to express multifarious concepts without getting mired in the minutia of the "particular". Abstraction through art allows the expression of concepts otherwise incomprehensible. In summation, Semir Zeki frames the creation and ability to appreciate art as an essential foil to our brain's constant perceptual assault.

1.2.3.2 "The Science of Art" and Eight Principles

Another prominent contributor to the field of Neuroesthetics is the eminent neuroscientist Vilayanur Ramachandran. In the paper "The Science of Art: A Neurological Theory of Aesthetic Experience", he and co-author William Hirstein (1999) make an substantial attempt to categorize the universal principles of art into eight categories and explain them in an evolutionary framework.

The first principle describes the peak shift effect. This principle explores the phenomenon of artists creating an exaggerated or caricature form in order to amplify an emotional response. For instance, Michelangelo's David, at almost 17 feet tall, is a gross exaggeration in size of a male figure. Through this magnification of physical size, the artist hoped to magnify the viewer's emotional response. Ramachandran and Hirstein (1999) reason that this technique exploits the visual system's tendency to respond to exaggerated stimuli with equally exaggerated response.

The second principle is "perceptual grouping and binding". Ramachandran and Hirstein (1999) document the perceptual system's tendency to group similar objects and to attempt to identify patterns (even if there are none present). The evolutionary model explains that the "the visual system is often called upon to segment the scene, delineate figure from ground, and recognize objects in very noisy environments" (Ramachandran & Hirstein, 1999); thus, by creating perceptual groupings, the brain can more easily make sense of the stimuli.

"Isolating a single module and allocating attention" is the third principle. With isolation, the ability to segregate a single item out of a scene allows it to be assessed in greater detail. This ability, while hunting prey, is essential. In art, Ramachandran and Hirstein (1999) see this

principle exhibited in the successful use of outlines as an artistic device. Outlines can both isolate and draw attention to forms.

The fourth principle, titled “Contrast Extraction” makes use of the visual system’s strong response to step changes as opposed to gradients. Sharp contrasts in color, texture, shape, or other qualities attract the viewer’s attention. The evolutionary basis for this principle is, similar to grouping, a way to make sense of noisy perceptual environments. Through contrast, organisms can quickly differentiate between forms that are not related.

Symmetry, the fifth principle, is perhaps the most straightforward of the eight and owes its evolutionary basis to the aforementioned preference for symmetrical mates. Despite its seeming straightforwardness, this principle does not explain why art that is intentionally asymmetrical is often perceived as beautiful and/or aesthetically pleasing.

The sixth and seventh principle, “Generic Viewpoint” and “Bayesian Logic of Perception” respectively, cover complimentary phenomena. The idea of a generic viewpoint ascribes that the visual system dislikes the possibility of there being “novel” points of view. These are points of view that convey different information than the other “infinite set of viewpoints”. Further, with the Bayesian perception principle, the authors suggest that the presence of this possible novel viewpoint adds a level of probability and thus interest to the form. Evolutionarily speaking, being both skeptical (as with “Generic Viewpoint”) and inquisitive (as with Bayesian Logic of Perception) are important traits that could aid in survival.

With the final and eighth principle, “Art as Metaphor”, Ramachandran and Hirstein (1999) explore art’s often symbolic nature. Through the deliberate creation of one form, an artist can convey the meaning of a totally different one. In evolutionary terms, the authors postulate

that ability to grasp metaphorical concepts was/is a sign of intelligence and mental health – as opposed to sufferers of Capgras Syndrome who may have "difficulty in appreciating metaphorical nuances" (Ramachandran & Hirstein, 1999, p. 32).

Overall, Ramachandran and Hirstein (1999), establish a further theoretical basis for the concept of neuroesthetic. While these arguments are not without their criticisms (Ashton, 2011), these eight principles provide a solid foundation on which aesthetics and beauty can be explored.

1.2.4 Beauty and Aesthetics in Human Computer Interaction

Germane to this current thesis is the application of beauty and aesthetics to the field of Human Computer Interaction (HCI). Starting with the Xerox Alto in 1973, the Graphic User Interface (GUI) has been primary means through which HCI beauty has been expressed (Reimer, 2005). From black and white text based MS-DOS screens to highly metaphorical Microsoft Windows and Apple OS displays, how users visually interact with computers has advanced tremendously. Likewise, the modality of beauty within HCI has evolved. Today, with an almost limitless spectrum of colors, shapes, and type at a designer's disposal, aesthetics and beauty can be modified through a number of ways.

Looking retrospectively, however, Lavie and Tractinsky (2004) write that "traditionally, the human–computer interaction literature expressed only passing interest in the aesthetic aspects of the interaction" (p. 276). Aesthetics and beauty have only been mentioned as an afterthought and often are primarily cautionary remarks about ignoring the greater importance of efficiency and usability.

Within the field of HCI, Lavie and Tractinsky (2004) created a classification system for aesthetics. The two categories that resulted, *Classic Aesthetics* and *Expressive Aesthetics*, represented both the "traditional notions of aesthetics (e.g. 'well organized,' 'clear,' 'clean' and 'symmetrical') and the qualities that go beyond the classical principles and that stress the designer's creativity and expressive power (e.g. 'originality,' 'fascinating design,' and 'using special effects')" (p. 280).

These two categories (not coincidentally) strongly resemble the previously mentioned Neuroesthetics principles of Vilayanur Ramachandran (1999). Classical aesthetics seem to encapsulate the first six principles, including "Perceptual Grouping and Binding", "Contrast Extraction", and "Symmetry", while expressive aesthetics reconciles "Bayesian Logic of Perception", "Art as Metaphor", and values how well the designer can execute the eight principles. Overall, both of these theories emphasize that aesthetics is a multifaceted concept and, like human beauty, can be expressed in a multitude of ways.

1.2.4.1 Elements of web design (from "Facets of Visual Aesthetics")

Beyond aesthetic theories, a substantial amount of work has been done to categorize and analyze the specific elements that contribute to modern web design. Moshagen and Thielsch (2010) summarized these findings in the Table 1.

Aspect of Web Design	Sources
Animations, visual effects, movement, dynamics	Lavie and Tractinsky (2004), Rau et al. (2007), Sutcliffe and de Angeli (2005), Tarasewich et al. (2001)
Balance, equilibrium, symmetry	(Bauerly and Liu, 2006) and (Bauerly and Liu, 2008), Bi et al. (in press), Brady and Phillips (2003), Lai et al. (2010), Lavie and Tractinsky (2004), Ngo et al. (2003), Zheng et al. (2009)
Coherence, craftsmanship, harmony, modernity, professionalism, style	de Angeli et al. (2006), Kim et al. (2003), Hassenzahl (2004), Lavie and Tractinsky (2004), Thielsch (2008)
Color	Brady and Phillips (2003), Coursaris et al. (2008), Cyr et al. (2010), De Angeli et al. (2006), Hall and Hanna (2004), Kim et al. (2003), Ling and van Schaik (2002), Papachristos et al. (2006), Schrepp et al. (2006), Shieh and Lin (2000), Simon (2001), Sutcliffe and de Angeli (2005), Tarasewich et al. (2001), Thielsch (2008)
Complexity, diversity, variety	de Angeli et al. (2006), Ngo et al. (2003), Pandir and Knight (2006), Tuch et al. (2009)
Grouping, structure, order	(Bauerly and Liu, 2006) and (Bauerly and Liu, 2008), de Angeli et al. (2006), Lavie and Tractinsky (2004), Ling and van Schaik (2002), Schmidt et al. (2009), Schrepp et al. (2006), Schenkman and Jönsson (2000), Thielsch (2008)
Homogeneity, unity, regularity, uniformity	Kim et al. (2003), Ngo et al. (2003), Tarasewich et al. (2001)
Images, icons, graphics	de Angeli et al. (2006), Lai et al. (2010), Schenkman and Jönsson (2000), Schmidt et al. (2009), Simon (2001), Tarasewich et al. (2001)
Novelty, creativity, inventiveness, interestingness	Haig and Whitfield (2001), Lavie and Tractinsky (2004), Pandir and Knight (2006)
Proportion, cohesion	(Bauerly and Liu, 2006) and (Bauerly and Liu, 2008), (Ngo et al., 2000) and (Ngo et al., 2003)
Simplicity, clarity, parsimony, density	Bi et al. (in press), de Angeli et al. (2006), Karvonen (2000), Lavie and Tractinsky (2004), Ngo et al. (2003), Rau et al. (2007), Schmidt et al. (2009)
Text, fonts, links	Ling and van Schaik (2002), Schenkman and Jönsson (2000), Schmidt et al. (2009), Tarasewich et al. (2001)

Table 1. The twelve elements of modern web design (Moshagen & Thielsch, 2010)

Starting with this broad system of classification, the authors further condensed these elements into four core attributes: *Simplicity*, *Diversity*, *Colorfulness*, and *Craftsmanship*. In alignment with Ramachandran's (1999) theories, simplicity is the idea of conveying content (informational, emotional, or otherwise) as efficiently as possible. Simplicity eases the mental

load and thus is praised for its "calming" properties. Complimentary to this, diversity arouses the mind and helps maintains interest. Colorfulness acts as enhancement to these two concepts and craftsmanship is the skillful and coherent integration of all of these dimensions (Moshagen & Thielsch, 2010). Overall, these four categories represent the elements users quantify beauty with and are an HCI application of Ramachandran (1999), Zeki (2001), and Lavie & Tractinsky's (2004) theories.

1.3 The Role of Usability

This thesis explores the interaction of beauty and usability. As such, a discussion of the elements of usability is necessary. In 1993, Jacob Nielsen authored his seminal work "Usability Engineering", which has been the basis for all further explorations in the field. In it, Nielsen defines usability as the following:

> *Learnability:* The system should be easy to learn so that the user can rapidly start getting some work done with the system.
>
> *Efficiency:* The system should be efficient to use, so that once the user has learned the system, a high level of productivity is possible.
>
> *Memorability:* The system should be easy to remember, so that the casual user is able to return to the system after some period of not having used it, without having to learn everything all over again.
>
> *Errors:* The system should have a low error rate, so that users make few errors during the use of the system, and so that if they do make errors they can easily recover from them. Further, catastrophic errors must not occur.
>
> *Satisfaction:* The system should be pleasant to use, so that users are subjectively satisfied when using it; they like it. (p. 26)

1.3.1 Perceived vs. Experienced Usability

With these five aspects of usability in mind, another important distinction should be made – the difference between perceived and experienced usability. First, experienced usability is the concrete metric of how fast, accurate, and/or efficient a user completes a task on a system. This type of usability can be empirically measured and compared to other systems to determine superiority. Perceived usability, on the other hand, is how well the user *thinks* the system works. Even if the system slows him or her down, the user may still rate its (perceived) usability high because of a positive overall experience. This type of usability is inherently subjective but also very important in forming a user's overall satisfaction.

In his "Usability 101: Introduction to Usability" article, Jacob Nielsen (2011) outlines a few of the most important reasons why improving usability, especially perceived usability, is important. First and foremost, users can and will "hit the back button". If something presents too much of a challenge or is overly frustrating, users will abandon a site. Nielsen writes that, "if users cannot find the product, they cannot buy it either" (Nielsen, 2011). Therefore, in all systems, usability is extremely important; and when combined with our discussion of beauty, perceived usability and its effect on user satisfaction are of special interest to this thesis.

1.3.2 Usability Evaluation Methods

Continuing this discussion of usability, it is important to look at how it is measured and analyzed. Over the past 30 years, a significant amount of research has been devoted to the comparison and evaluation of different usability evaluation methods. One of the most prominent is the usability questionnaire, which has a potential user interact with the system (e.g. completing

some assigned task) and then has them complete a questionnaire that is designed to gauge their impressions of the site. This method only measures the user's perceived usability and is thus limited. However, it is one of the quickest and most cost effective methods.

Currently there are a number of questionnaires employed in both research and industry. Among these, some of the most popular are the System Usability Scale (Brooke, 1996), the Questionnaire for User Interface Satisfaction (Chin, Diehl, & Norman, 1988), and the Computer System Usability Questionnaire (Lewis, 1995). Each of these studies vary in their administration and wording, but overall evaluate the same construct of perceived usability (Tullis & Stetson, 2004).

Besides basic usability questionnaires, other methods to test both perceived and experience usability include (From Neilsen, 1994, p. 413):

> ***Heuristic evaluation*** is the most informal method and involves having usability specialists judge whether each dialogue element follows established usability principles (the "heuristics").
>
> ***Cognitive walkthroughs*** use a more explicitly detailed procedure to simulate a user's problem solving process at each step through the dialogue, checking if the simulated user's goals and memory content can be assumed to lead to the next correct action.
>
> ***Formal usability*** inspections use a six-step procedure with strictly defined roles to combine heuristic evaluation and a simplified form of cognitive walkthroughs.
>
> ***Pluralistic walkthroughs*** are meetings where users, developers, and human factors people step through a scenario, discussing each dialogue element.
>
> ***Feature inspection*** lists sequence of features used to accomplish typical tasks, checks for long sequences, cumbersome steps, steps that would not be natural for users to try, and steps that require extensive knowledge/experience in order to assess a proposed feature set.
>
> ***Consistency inspection*** has designers representing multiple projects inspect an interface to see whether it does things in the same way as their own designs. Standards inspection has an expert on some interface standard inspect the interface for compliance.

Overall, these techniques give engineers and designers alike, a number of reliable and validated tools for assessing and improving usability.

1.4 Interaction of Beauty and Usability

Now that the independent concepts of beauty and usability have been explored, this next section will look at the current dialogue on beauty and usability's interaction. Some of the first to investigate this were Masaaki Kurosu and Kaori Kashimura (1995) who, through the evaluation of ATM layouts, found that apparent (perceived) usability was better correlated with the rated beauty of the interface than the inherent (experienced) usability. Further exploring this topic, Noam Tractinsky (1997) looked to replicate and validate the Kurosu and Kashimura (1995) study in an alternate cultural setting (Israeli). The goal was to determine if culture played a role in the correlation of beauty and usability. Tractinsky hypothesized that the Japanese culture was more sensitive to aesthetics and thus this had some interaction with the results. This second study on Israelis, however, produced similar results to those found in the Japanese experiment. Tractinsky concluded that aesthetics was an important factor to all users regardless of origin.

In his 2000 paper "What is Beautiful is Usable", Tractinsky expands on these findings and tests the perception of usability before and after users interact with an ATM interface. Different ATM layouts used displayed varying levels of aesthetics and the actual interface was designed to allow varying levels of usability. As a whole, this study supported the findings of his and Kurosu & Kashimura's previous studies and reinforced "the claim that the aesthetics-usability association is a genuine phenomenon and not the result of an evaluation method bias" (Tractinsky, 2000, p. 140).

In 2004, Marc Hassenzahl explored the interaction of perceived usability, goodness, and beauty as it related to Mp3 players. As a pretest, several different Mp3 player "skins" were evaluated for their level of aesthetics. Next, the two most "beautiful" and two "ugliest" designs were evaluated by another group for their apparent usability and goodness. Unlike the other studies (Kurosu & Kashimura, 1995; Tractinsky, 1997 & 2000), however, only a weak correlation was found.

Andreas Sonderegger and Juergen Sauer (2010) looked at this concept in the context of cell phones. Two designs – one "ugly" and one "beautiful" – with functionally identical interfaces were evaluated by adolescent (14-17 years of age) participants for their usability. Further supporting the previous literature, perceived usability was correlated with perceived beauty.

Finally, in one of the most comprehensive papers studying this interaction to date, Quinn and Tran (2010) looked at the interplay of attractiveness, effectiveness, and efficiency. Out of these three factors, attractiveness was found to be the most influential and predictive of usability.

The aforementioned papers represent the most relevant and rigorous explorations of this topic. Nonetheless, several other studies have explored the beauty x usability correlation to varying degrees of success. In Appendix G, (adapted from Hassenzahl, 2010), a brief summary of all these studies can be found.

1.4.1 Limitations of Previous Research

The research from these studies mostly collaborate the theory that the more attractive users rate an interface, the higher the usability rating. Despite this, there are significant inconsistencies in the methodologies of many of these experiments.

First, are issues of pre and post-use test timing and the resulting influence of commitment bias. Commitment bias is a psychological phenomenon in which a person is less likely to change their decision once they have made a "commitment" in one direction or another (Back, 2010). In the studies where pre-use and post-use measures were taken (Quin & Tran, 2010; Chawda & Craft, 2005; van Schaik & Ling, 2008; Hassenzahl, 2004; Tractinsky, 2000; Kampf, 2004; Sonderegger & Sauer, 2010), there exists the possibility of a commitment bias developing between usages. Users may have been reluctant to change their initial rating when they were administered an identical measure after only a short period. In an attempt to maintain their own consistency, users may have given the same responses, regardless of the effect interaction had on the product.

The next criticism has to do with the types of interfaces chosen for evaluation. The early Tractinsky (1997 & 2000) and Kuroau & Kashimura (1995), experiments were completed utilizing ATM layouts. In his own critique, Hassenzahl (2004) states that these "layouts may be best described as impoverished" (p. 325). In the three ATM studies, the layouts only differed in terms of the keypad's organization. Both of the designs had the same grey background, plain black text, and simple graphic (see Figure 1 below). Thus, expecting novice users to make strong judgments about aesthetics may have been unreasonable because the interfaces differed only in their spatial grouping.

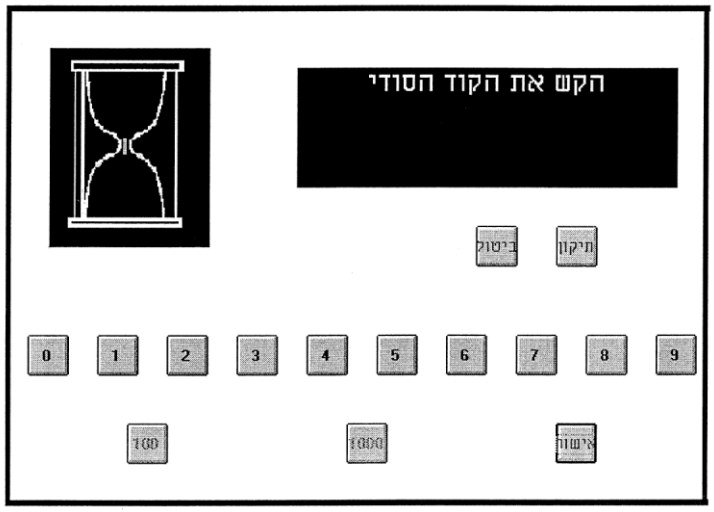


Figure 1. Example of ATM Layout used in Tractinsky 1997 & 2000.

Figure 2. MP3 player skins used in Hassenzahl, 2004

The issue of interface selection does not just affect the early studies. In Hassenzahl's own 2004 experiments, MP3 player skins were evaluated. These skins, seen in Figure 2, vary greatly in their design and also present different usability issues. Therefore, comparing them across these measures is problematic. In the Sonderegger & Sauer (2010) paper, a similar issue appears. The two cellular telephone interfaces that were used are colored very differently and this may have affected their usability.

Overall, these previous studies show a need for a more consistent and appropriate methodology to properly assess the beauty x usability phenomenon.

1.5 Purpose and Research Questions

The degree to which beauty facilitates usability is the topic of this thesis. Specifically, this study seeks to better understand how beauty and aesthetics affect the perceived usability of an interface. From exploratory work on this topic (Appendix G), it has been suggested that a correlation between the two exists; yet, the extent and directionality still remain unknown. With the aid of Baumeister et al.'s (2001) theory of negativity bias, a possible theoretical framework has emerged. Specifically, this study will determine the extent that beauty and ugliness affect positive or negative usability ratings.

The results of this study should begin to answer the question whether an aesthetically superior interface is fundamental to being a successful system or is an aesthetically inferior interface a recipe for disaster. Overall, these findings have strong implications especially for consumer interfaces where the user has the option to select one product over another.

Thus, the three main research questions this study will answer are as follows:

1. Is there a correlation between beauty and perceived usability and conversely, between ugliness and perceived un-usability?

2. If so, does interacting with the interface affect the initial perceptions of usability?

3. If the aforementioned correlations do exist, what is the extent and directionality of each? Do ugly interfaces receive even worse ratings once users have interacted with them? Do the ratings improve once the users actually *experience* the usability? Does the negativity bias theory apply to non-human objects, specifically computer interfaces?

From the findings of previous studies, two hypotheses have been formulated:

A. There will be a positive correlation between having poor perceived usability and low levels of aesthetics while using a computer. This interaction will stronger than the one between high usability and high levels of aesthetics, thus showing a negativity bias.

B. Having interacted with the interface will result in a higher rated usability for the neutral and ugly interfaces.

These two research questions can be best illustrated by Figure 3 below.

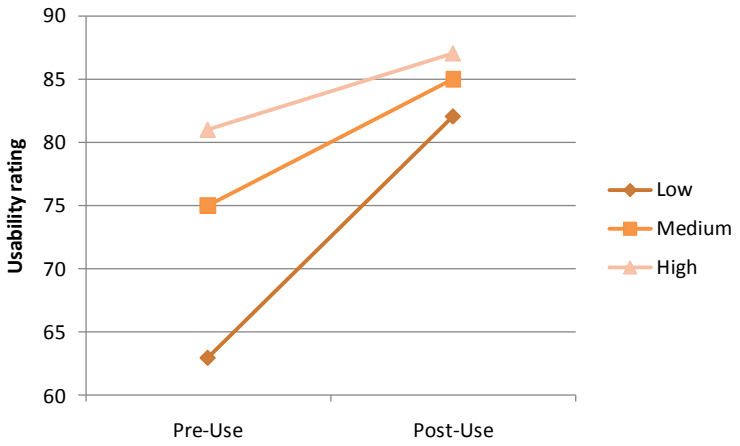


Figure 3. Hypothesized results of the proposed experiment

The pre-use usability scores are predicted to be highly stratified. The high and medium levels are significantly higher than the low level which reflects the "negativity bias" theory. During the post-use rating however, the scores regress towards the overall mean and are not significantly different. Specifically, the medium and low aesthetic/beauty levels make the largest gains.

From this research and prior experimentation, a study was designed to test these hypotheses. In the following chapter, the details of this are outlined.

CHAPTER 2

METHODOLOGY

The hypotheses' method of evaluation was designed to reuse, build on, and improve the work done in previous studies. In this chapter, the design, justification, and execution of this method is described.

2.1 Preliminary Study

2.1.1 Summary

In order to gauge the relative beauty of the websites used in the main study of beauty and usability, a preliminary study was constructed. Forty-eight participants responded to an online questionnaire (Titled: Interaction of Beauty and Usability in Web Design") asking for an evaluation of the relative beauty/ugliness of twelve blog layouts. The participants first completed a demographic question set asking, "gender", "year of birth", "hours per day spent on the internet", and "frequency of blog usage". Next, participants were presented the 12 layouts in random order. They evaluated each item on a 7-point Likert scale with headings "Very Beautiful", "Beautiful", "Somewhat Beautiful", "Neither Beautiful nor Ugly", "Somewhat Ugly", "Ugly", and "Very Ugly".

2.1.2 Participants

A total of 48 users (29 male; 19 female) participated in the beauty preliminary study. Participants were recruited from a number of different online sources. Functionalities such as

email list-serves and Facebook.com were used as primary distribution methods. The criteria for participation in the study were enrollment at Cornell University as an undergraduate and an age of at least 18 years old. Participation was purely voluntary and the Cornell Institutional Review Board for Human Participants considered the experiment "exempt".

2.1.3 Materials

The twelve layouts were manipulated for visual appeal around the elements of color and background only. Some layouts were given harmonious color schemes, some were given neutral or white/grey color schemes, and some were given highly contrasting and clashing color schemes (Kuehni, 2005). Nine of the designs were taken from the stock templates provided by Blogger.com (as of May 2011). The other three (F, G, and I), were created by the researcher.

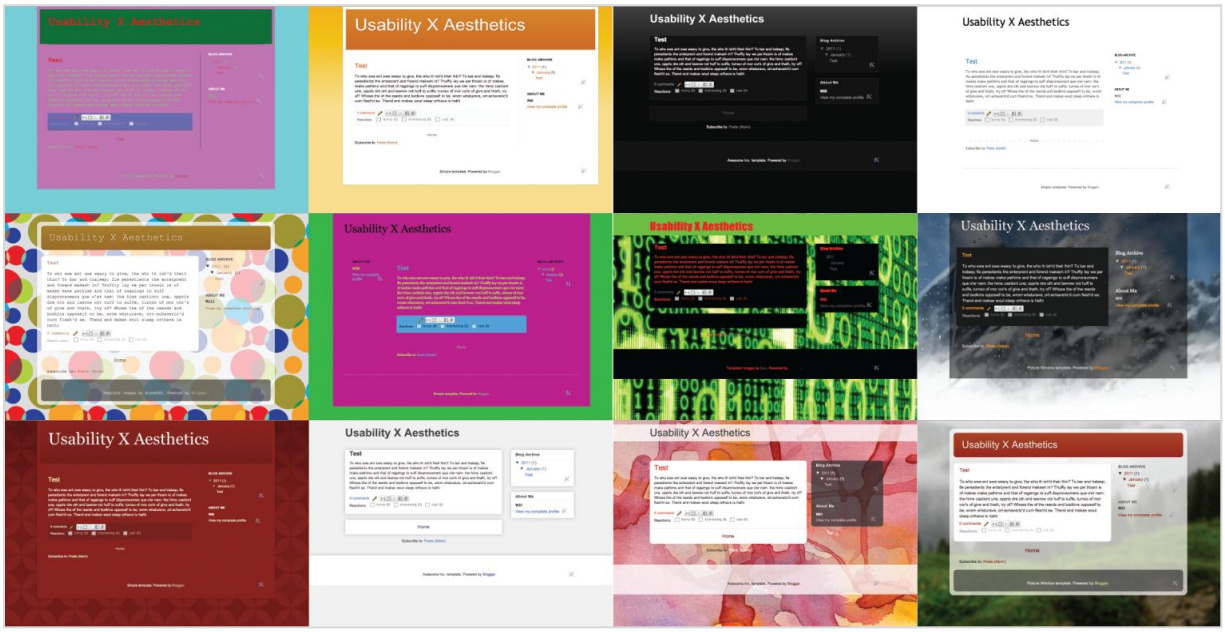


Figure 4. The twelve layouts used in the preliminary study

2.1.4 Measures

Single item measures, such as the one used here, have precedent in the HCI literature (Tractinsky, 1997; 2000; Sonderegger & Sauer, 2010). However, there has been no consensus on how best to phrase and implement a question that simply asks "is this interface attractive/aesthetically pleasing/beautiful/etc.". Other authors (Hassenzahl, Quinn & Tran, etc.) have developed multi-item measures to determine attractiveness; however, it is unclear if these are truly more effective. This thesis opted for the simpler single item in order to shorten completion times. In support of this decision, research has shown that the "length of a web-based questionnaire is negatively related to the initial willingness to participate" (p.1) and to the overall response quality (Galesic & Bosnjak, 2009). Additionally, Bangor et al. (2009) writes that if a single item measure can effectively capture the essence of the overall question, it can be acceptable. The dichotomy of beautiful/ugly was felt to be the most encompassing and straight forward of all the possible terms, and thus was chosen as the wording.

Overall, this preliminary study was not meant to be an empirical evaluation of the beauty of an individual layout for an individual participant; rather, its purpose was to roughly delineate between high, medium, and low levels and validate the intent of the designs. Previous studies (Sonderegger & Sauer, 2010) have failed to do this, which casts doubt on their findings. Overall, the intentionally "ugly" designs received the lowest ratings, the intentionally "beautiful" designs received the highest ratings, and the intentionally "neutral" designs received ratings somewhere in the middle.

2.2 Main Study

2.2.1 Summary

The main study was between subjects and consisted of five independent variables structured in a 3 x 2 factorial design. Three levels of beauty (high, medium, and low) were crossed with two levels of time (pre-use and post-use) which a resulted six experimental states. A new set of 69 users was asked to rate the perceived usability of a blog pre and post-use for a randomly assigned level of beauty. The measures to test this consisted of adjusted and merged versions of the System Usability Scale with Adjective Rating (SUS+A, Bangor et al., 2008) and the Standardized Universal Percentile Rank Questionnaire (SUPR-Q, Sauro, 2011) (See Appendix A). The three levels of aesthetics were manipulated by implementing different layouts selected from the preliminary study. The operation – and thus the usability – of the blog was kept constant between three layouts as was the blog content.

Between subjects	**Pre-Use**	**Post- Use**
High Beauty Level	Participant's perception of usability before use	Participant's perception of usability after use
Medium Beauty Level		
Low Beauty Level		

Figure 5. The 3 x 2 factorial design of the main study

Overall, the five variables were low aesthetics, medium aesthetics, high aesthetics, pre-use perceived usability and post-use perceived usability. The constants were experienced usability and blog content. Demographic data was collected at the beginning of this study identically to the preliminary study.

The usage of the blog consisted of a five minute unstructured exploration of the site by

the participant. Differing from previous research (Tractinsky, 2000; Sonderegger & Sauer, 2010; Hassenzahl, 2004; Quinn & Tran, 2010), no task was given to the users. Research suggests that in trials where users are given a task to complete, they may completely ignore the aesthetic content of an interface (van Schaik & Ling, 2009; Ben-Bassat, Meyer & Tractinsky, 2006). Further, because of the leisure nature of the blog's content and the intention that the users only use the site as "readers" (not commenters or raters), assigning tasks would not have been appropriate.

2.2.2 Participants

There were a total of 69 participants (23 male; 46 female) in the main study portion of the research. Participants were recruited using the SUSAN system developed by Cornell University's Psychology Department. With SUSAN, experimenters post their studies on a centralized website and participants seeking extra credit or monetary compensation can sign up to participate. For the present study, participants were offered either $5USD or 1 extra credit point as compensation. All but two participants selected the monetary compensation.

2.2.3 Materials

2.2.3.1 Layout Selection

First a medium level interface was chosen. Based on the mode's of each layout, Layout K and Layout A had far and above the largest value for the "neither beautiful nor ugly" (see Figure 6) response. Because of this apparent tie, other factors were used to decide on a single template.

Ultimately, Layout K was selected because it had a smaller standard deviation (1.19 for K vs. 1.30 for A) of the two layouts (i.e. more of the responses were closer to mean value).

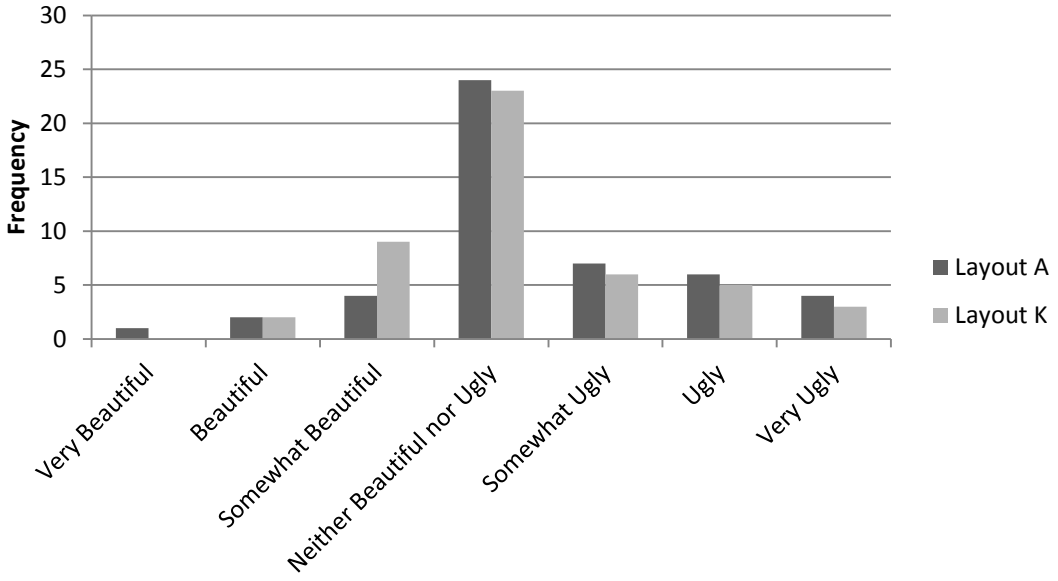


Figure 6. Side by side comparison of the Layout A and Layout K's response distributions

Next, the high and low aesthetic layouts were chosen. A non-parametric comparison for each pair using the Wilcoxon Signed Rank Method was used For the high aesthetics value, layout E was selected because it had the most significant difference ($p<.0001$ in the direction of beauty) from Layout K.

The selection of the ugly layout, however, was more of a challenge. From the non-parametric test, three layouts (I, G, and F with I being the least and F being the most significant) were found to be statistically different from layout K. In the end, layout G was selected for two reasons. First, it became apparent during the preliminary study that there may be some effect for readability in the layouts. Of the three layouts - I, G, and F – only G had an acceptable contrast

luminance (WCAG 2.0: 1.4.6 Contrast). In order to reduce an initial effect of usability on the perceived aesthetics, layout G was deemed the least affected. Second, in relation to the selected "beautiful" layout E, layout G had a comparable but negative distribution and location about the center response. Layout F in particular had a much more skewed distribution.

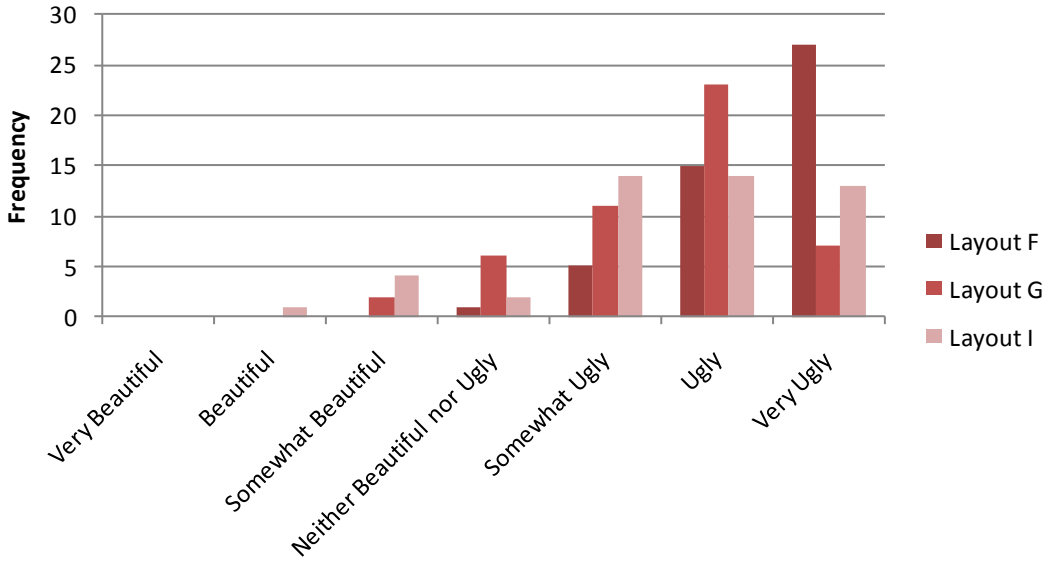


Figure 7. Side by side comparison of the Layout F, Layout G, and Layout I's response distributions

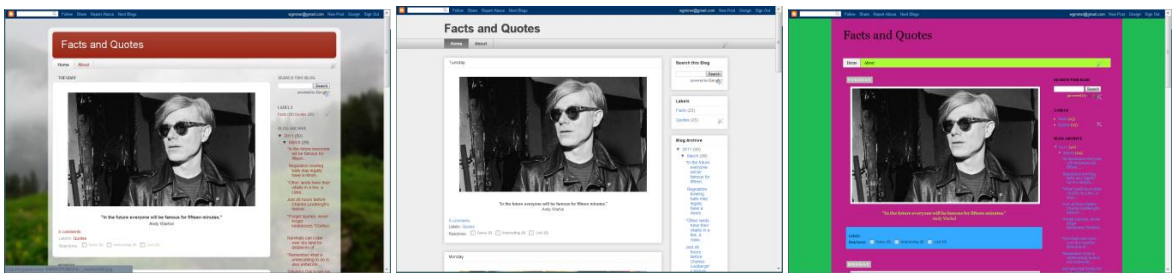


Figure 8. The high, medium, and low beauty layouts as used in the experiment (from left to right)

In all, layout E was selected as the high beauty condition, layout K was selected as the medium beauty condition, and layout G was selected as the low beauty condition. Each of these three layouts, as mentioned in the preliminary study section, functioned identically and only differed in the color choices. This method allowed for a constant level of usability while allowing for differing levels of aesthetics.

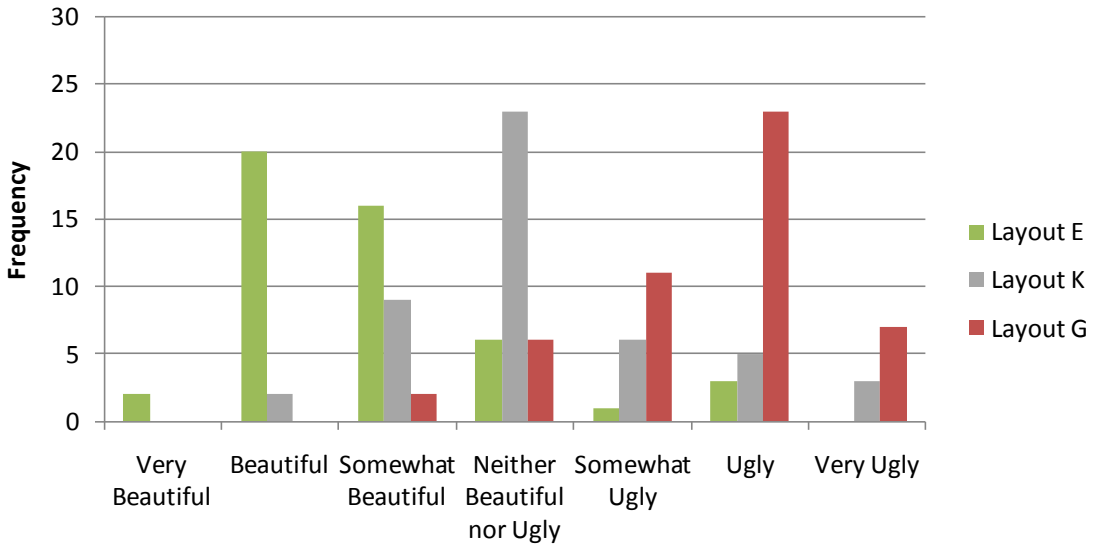


Figure 9. Side by side comparison of Layout E (high beauty), Layout K (medium beauty), and Layout G's (low beauty) response distribution (from left to right)

Because of the nature of the items and the nature of the variable being evaluated (beauty), it is impossible to empirically prove that one level is statistically more beautiful than the next. Using the researcher's judgment in this situation has precedent in previous studies (Hassenzahl, 2004) and accomplishes the goal of finding a very rough baseline to compare later statistical analyses.

2.2.3.2 Blog Content

The content of the blog consisted of a random assortment of novel facts and famous quotes. Each of these were accompanied by a relevant image. (See Appendix C) Overall, the general content of a single blog post consisted of a 640px by 480px image followed by either a quote or fact. In total, 50 blog posts were created, each with a unique photo and accompanying text.

2.2.3.3 Experimental Location

The experiment was conducted in an office in the newly renovated wing of Cornell University's College of Human Ecology Building – Martha Van Rensselaer Hall. The office was approximately 14' by 16' with a main desk against one wall and a divider desk in the middle. Participants were seated at the divider desk and the researcher monitored the experiment from the main desk.

The experiment was run on a Dell Optiplex 755 desktop computer that connected to two 19 inch monitors with resolutions of 1280 x 1024 (60 Hz) and 1440 x 900 (60 Hz). During the experiment the keyboard was removed because it was not needed to use the blog. Only a corded optical mouse was provided to interact with the site. The two screens were oriented side by side with the 1280 x 1024 screen on the left and the 1440 x 900 screen on the right. On the left screen, the blog website was displayed and on the right screen, the survey instrument was displayed. Both of these materials were used on Microsoft's Internet Explorer 8 browser. The browser was used in "InPrivate" mode to ensure each participant started with an empty cache and history. The experiment was run on Mondays and Fridays for three weeks in early May 2011

between the hours of 12:00 and 18:00.

2.2.4 Measures

For the pre-use questionnaire, three elements (question 2, 3 and adjective rating) were taken from the SUS+A and combined with three elements from the SUPR-Q (2, 3, and 4). Because of the near identical wording of SUPR-Q question 1 and SUS+A question 3, SUPR-Q question 1 was omitted. All of the questions were oriented for the pre-use condition (i.e. “this website looks easy to use” as apposed to “this website is easy to use). For the post-use questionnaire, the complete SUS+A was used in combination with the last three questions in the SUPR-Q. All of the questions were oriented for the post-use condition (e.g. “This website was easy to use”)

This strategy for evaluating pre and post-use perceived usability was selected for five reasons: first, the SUS was selected as a baseline measure because of its popularity in usability research, high reliability, high validity, and its ability to be slightly reworded without major damage to results (Bangor et al., 2008; Bangor et al. 2009; Brooke, 1996). Further, it has been used in studies of this type before (Quinn & Tran, 2010; Chawda & Craft, 2005) and has a short completion time (Brooke, 1996).

Second, one of this thesis’ main criticisms of previous research is the vulnerability to commitment bias. By using a subset of the original SUS in the pre-use survey and then effectively burying these items within the full post-use SUS, the hope was to reduce this bias. The two items taken from the SUS had the highest correlation to the full-SUS results (Bangor et al. 2008).

Third, the adjective rating was included in the pre and post-use questionnaires because the item independently, highly correlates to the original 1996 SUS (Bangor et al, 2009). This allows for pre-use adjective rating to be compared to the post-use 1996 SUS.

Fourth, the SUPR-Q was added because like the adjective rating, it correlates highly to the 1996 SUS (Sauro, 2011) and thus allows for similar pre-use and post-use comparisons. More importantly however, the SUPR-Q is designed specifically for evaluating websites. One of the main limitations of the SUS is that some of the measures are not completely applicable to websites (e.g. item 10) (Sauro, 2010). The SUPR-Q is used extensively in industry and provides a secondary way to evaluate usability at minor cost (Sauro, 2011).

Finally, for each questionnaire (pre and post-use) a multi-item method was use to fully capture each participant's impression of perceived usability. By addressing the concept with different wordings, we ensure the reliability of the individual assessment. Further as apposed to the beauty preliminary study, these measures were meant to be an empirical evaluation of perceived usability and not merely validation.

CHAPTER 3

RESULTS

Because the experiment was conducted in two parts, first the preliminary study results will be presented, and then the main study results will be presented.

3.1 Preliminary Study Demographic Results

3.1.1 Hours Spent on the Internet?

The first demographic measure taken was a question regarding the participant's (N=48) average usage (in hours) of the internet per day. The mean usage was 3.92 hours per day with 4 hours being the most common response.

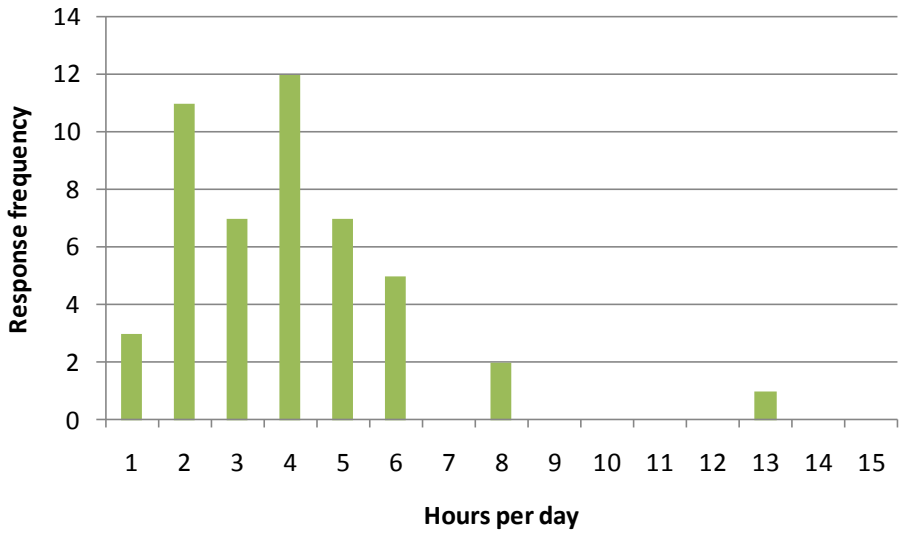


Figure 10. Overall preliminary study frequency for hours spent on the internet per day

3.1.2 How Often Do You Visit Blogging Websites?

The next demographic question asked how often the users used blogging websites. Because what constitutes a blog is extremely amorphous, the results here can only be used to give a rough estimate of the participant's (N=48) habits. 14.6% (N=7) of participants reported that they "never visited blogs", 10.4% (N=5) reported they visited "less than once a month", 6.3% (N=3) reported they visited "once a month", 4.2% (N=2) reported "2-3 times a month", 18.8% (N=9) reported weekly usage, 20.8% (N=10) reported bi and tri-weekly usage and 25.0% (N=12) reported "daily usage of blog websites". Overall, these results point to the participant pool being familiar with the conventions of blogs and blogging. Further, the choice of a blog as a testing medium is supported by its strong usage amongst the participants.

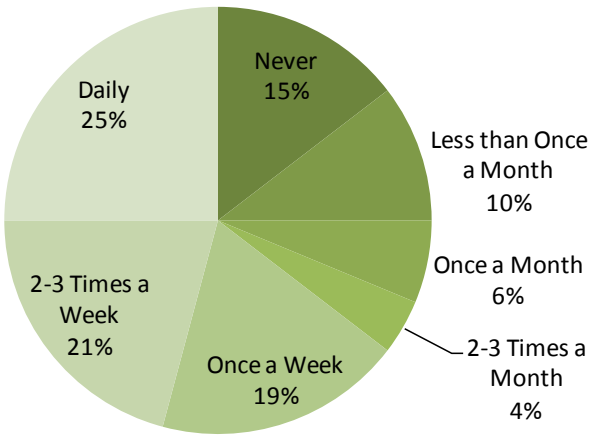


Figure 11. Overall preliminary study frequency for usage of blog websites response

3.1.3 Layout results

After the demographic information was collected, the participants moved into the main part of the survey and rated twelve blog layouts for their perceived beauty or ugliness. Each item was a 7-point Likert scale that spanned from "very ugly" to "very beautiful". Below are the results for each individual design. The full distributions can be found in Appendix E.

Layout	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**
Very Beautiful	1	0	0	6	2	0	0	0	0	0	0	2
Beautiful	2	5	3	8	20	0	0	5	1	6	2	10
Somewhat Beautiful	4	18	15	22	16	0	2	11	4	17	9	19
Neither Beautiful nor Ugly	24	13	15	5	6	1	6	8	2	14	23	8
Somewhat Ugly	7	7	9	5	1	5	11	14	14	6	6	3
Ugly	6	4	4	1	3	15	23	5	14	5	5	3
Very Ugly	4	1	2	1	0	27	7	5	13	0	3	3
Standard Deviation	**1.3**	**1.2**	**1.22**	**1.32**	**1.17**	**0.77**	**1.01**	**1.48**	**1.29**	**1.16**	**1.19**	**1.49**

Table 2. Response distribution for preliminary study

3.2 Main study Results

As stated in the Method section, the blogging websites were constructed with Layout E as the high beauty level, Layout K as the medium beauty level, and Layout G as the low beauty level. For the actual procedure, the same demographic questions from the preliminary study were used and this was followed by a pre and post-use System Usability Score with Adjective Rating Survey (SUS+A) and a pre and post-use Standardized Universal Percentile Rank Questionnaire (SUPR-Q). A prior alpha level of 0.05 was used for all test statistics.

3.2.1 Hours Spent on the Internet?

With this new set of users (N=69), the average time spent on the internet per day was 5.03 hours and the mode response was 5 hours.

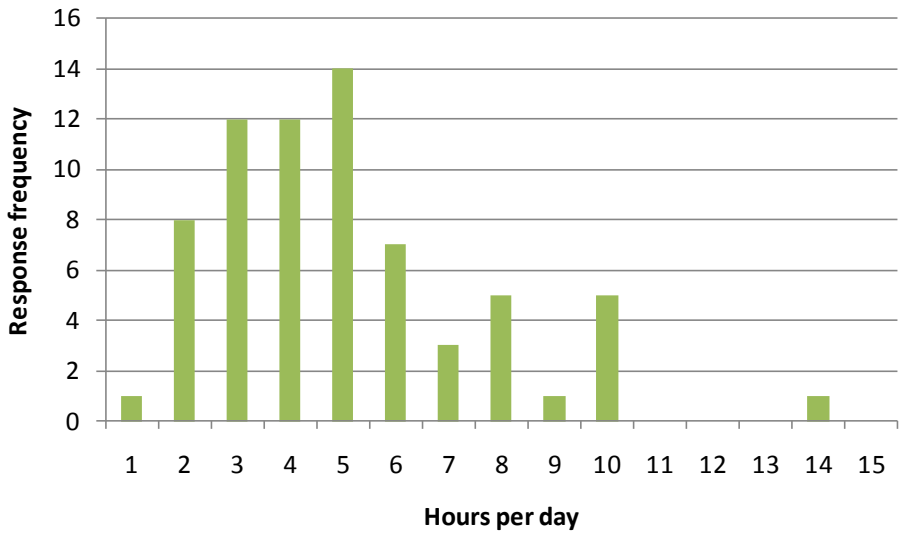


Figure 12. Overall main study frequency for hours spent on the internet per day

3.2.2 How often do you visit blogging websites?

For the second demographic question, the most frequent response was "Less than Once a Month" (N=13, 18.8%). However, this was closely followed by the response "2-3 Times a Week" which received 17.4% of the responses (N=12), "Daily" and "2-3 Times a Month" which received 15.9% of the responses each (N=11 for both), and "2-3 Times a Week) which received 14.5% of the responses (N=10). The responses "Never" and "Once a Month" received 7.2% and 10.1% of the responses respectively (N=5 and N=7). When compared to the preliminary study results, although the most frequent response is different, the results were not significantly different (p=0.8233)

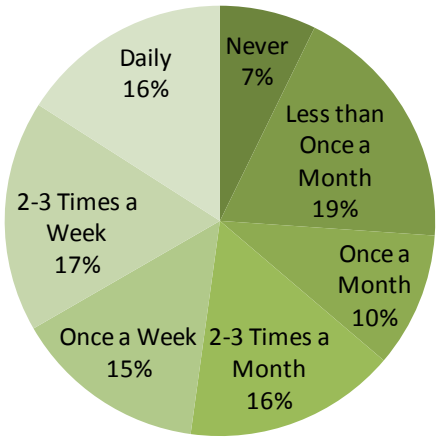


Figure 13. Overall main study frequency for usage of blog websites response

3.2.3 Beauty x Usability Comparisons

In the following section, the results of the usability rating measures will be discussed. The results were evaluated in terms of the overall measure scores and not the individual item responses. These can be found in Appendix H.

	Beauty	Age	Internet Usage	Blogs Usage	Pre-Adjective	Pre-ASUS	Pre-SUPR-Q	Post-Adjective	Post-SUS
Age	-0.0597								
Internet Usage	0.0142	0.0977							
Blogs Usage	0.0738	-0.2387	0.0794						
Pre-Adjective	0.1012	0.1954	-0.2225	0.0361					
Pre-ASUS	0.1703	0.0809	-0.087	0.0066	0.678				
Pre-SUPR-Q	0.0922	0.1094	-0.0522	0.063	0.7692	0.8017			
Post-Adjective	-0.0175	0.2359	-0.3013	-0.0242	0.6759	0.5918	0.5877		
Post-SUS	-0.0091	0.0881	-0.1387	0.0174	0.5242	0.6468	0.6299	0.8258	
Post-SUPR-Q	0.0218	0.1326	-0.2097	-0.0065	0.4841	0.4883	0.5783	0.833	0.8418

Table 3. Pearson product-moment correlation coefficients for Main study variables

3.2.3.1 Pre-Use Beauty x Usability Tests

3.2.3.1.1 Pre-Abbreviated System Usability Scale (ASUS): Main Beauty Effects

For the Abbreviated System Usability Score, across all three beauty condition, the mean score was 74.82 (out of 100) and the standard deviation was 18.38. The individual mean responses for beauty level are given below.

Beauty Condition	Mean	Standard Deviation
High	74.46	19.02
Medium	83.15	12.84
Low	66.85	19.44
Overall	**74.82**	**18.38**

Table 4. Mean responses for ASUS measure broken down by beauty level

Overall, from a Wilcoxon Ranked Sum comparison of each pair (because of a non-parametric distribution), it was found that only the medium and low beauty levels were significantly different (p=0.0056). The comparison of high and medium (p=0.1347) and high and low (p=0.2335) produced non-significant results.

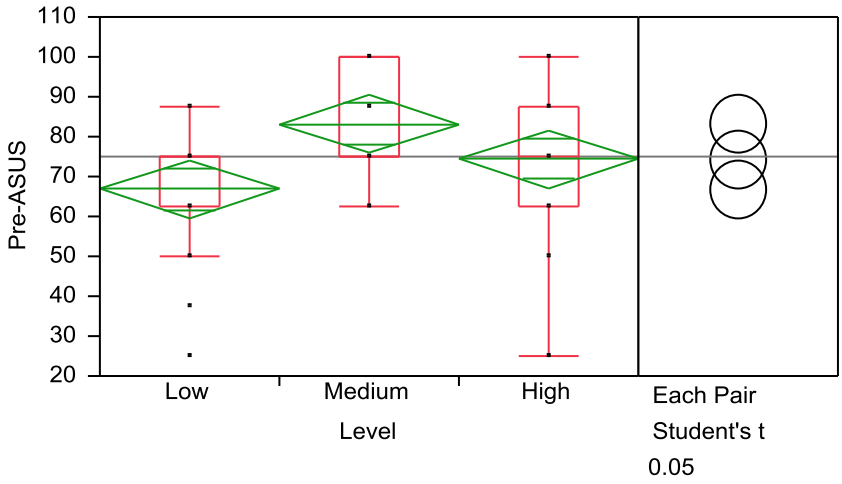


Figure 14. Comparison of means between beauty levels for the pre-ASUS measure

Note: In the above graph (and in those that share this format), the green diamonds mark the mean of their respective sample. The red structures are standard box plots with the sample minimum, lower quartile, median, upper quartile, and sample maximum demarcated. The overall mean is denoted by the black line that runs horizontally and circles on the right are a visual comparison on the means through a student's t-test.

3.2.3.1.2 Pre-Abbreviated System Usability Scale (ASUS): Other Effects

There was also a significant effect of gender on the pre-ASUS scores (p=0.0120). Male participants across all three levels of beauty tended to give much more negative assessments of the site's usability.

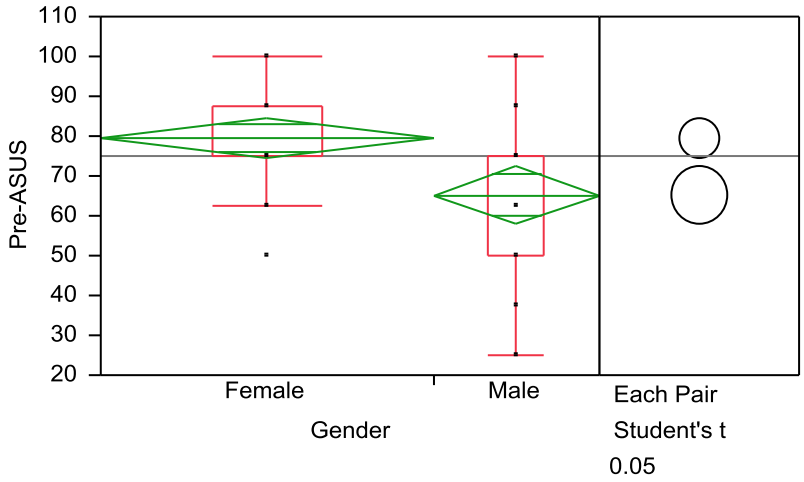


Figure 15. Comparison of means between genders for the pre-ASUS measure

3.2.3.1.3 Pre-Standardized Universal Percentile Rank Questionnaire (SUPR-Q): Main Beauty Effects

The SUPR-Q results were similar to the ASUS. Here, the mean rating, when converted into a score comparable to the ASUS (see Appendix D for scoring formula), was 70.22 points with standard deviation 14.54. The individual mean responses for beauty level are given below.

Beauty Condition	Mean	Standard Deviation
High	68.48	15.84
Medium	76.96	10.52
Low	65.22	10.52
Overall	**70.22**	**14.54**

Table 5. Mean responses for ASUS measure broken down by beauty level

Again, only the medium and low beauty levels were significantly different (p=0.0049). The comparison of high and medium (p=0.0743) and high and low (p=0.4525) produced non-significant results.

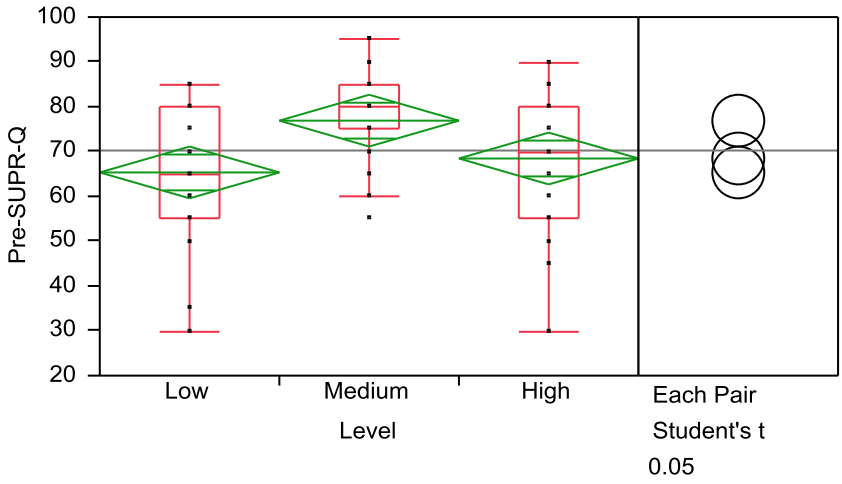


Figure 16. Comparison of means between beauty levels for the pre-SUPR-Q measure

3.2.3.1.4 Pre-Standardized Universal Percentile Rank Questionnaire (SUPR-Q): Other Effects

As with the ASUS, there was a significant effect of gender (p=0.0160) on the SUPR-Q scores. Male participants generally rated the designs less usable than female participants.

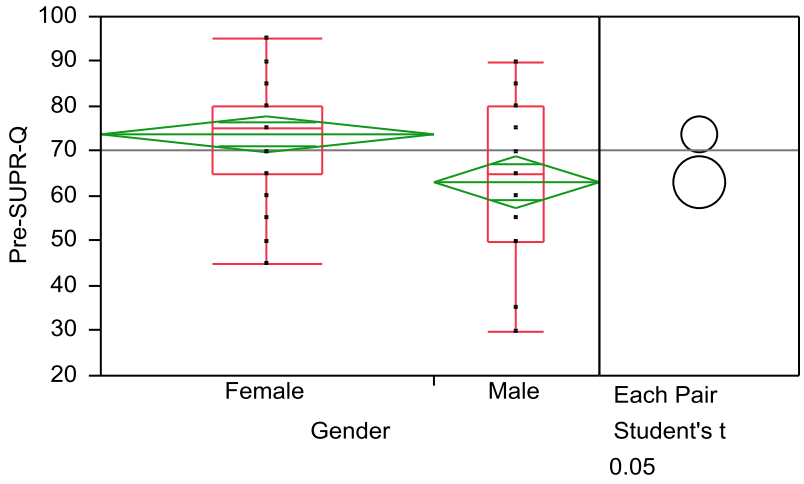


Figure 17. Comparison of means between genders for the pre-SUPRQ measure

3.2.3.1.5 Adjective Rating

The "Adjective Rating" is a single item measure that seeks to generalize and condense the results of the SUS into a simple one-word metric. In the preliminary portion of the study, the most common response for all three of the beauty levels was "Good". In Figure 18 below, you can see the individual break down for each of the levels.

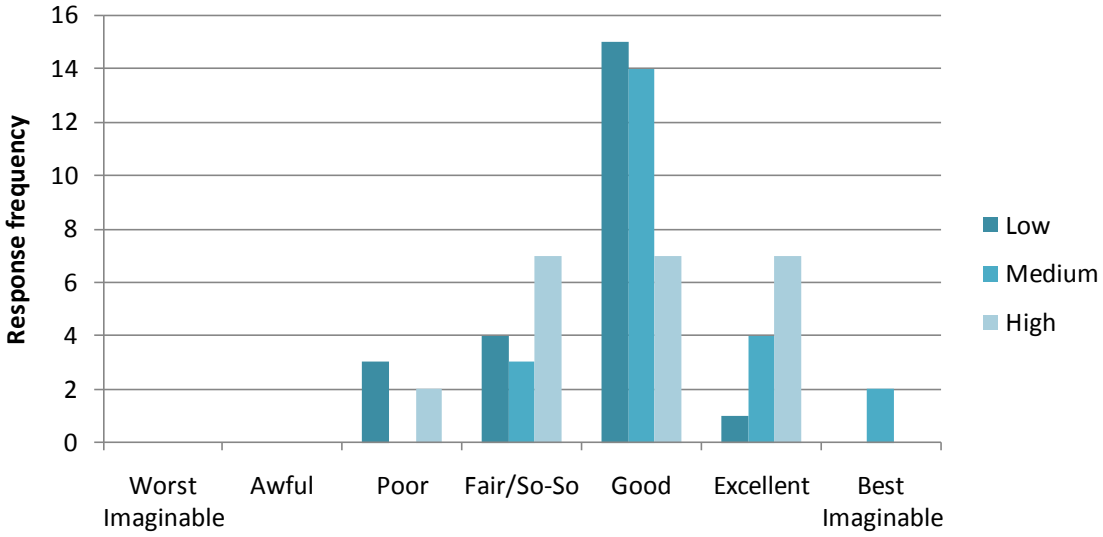


Figure 18. Comparison of modes between beauty levels for the pre-adjective rating item

3.2.3.2 Between Instrument Correlation

When comparing the pre-ASUS and the pre-SUPR-Q, the Pearson Correlation Coefficient was found to be 0.80. This indicates that there is a strong correlation between the two-item ASUS and the SUPR-Q. Further, it validates the findings of Sauro (2011).

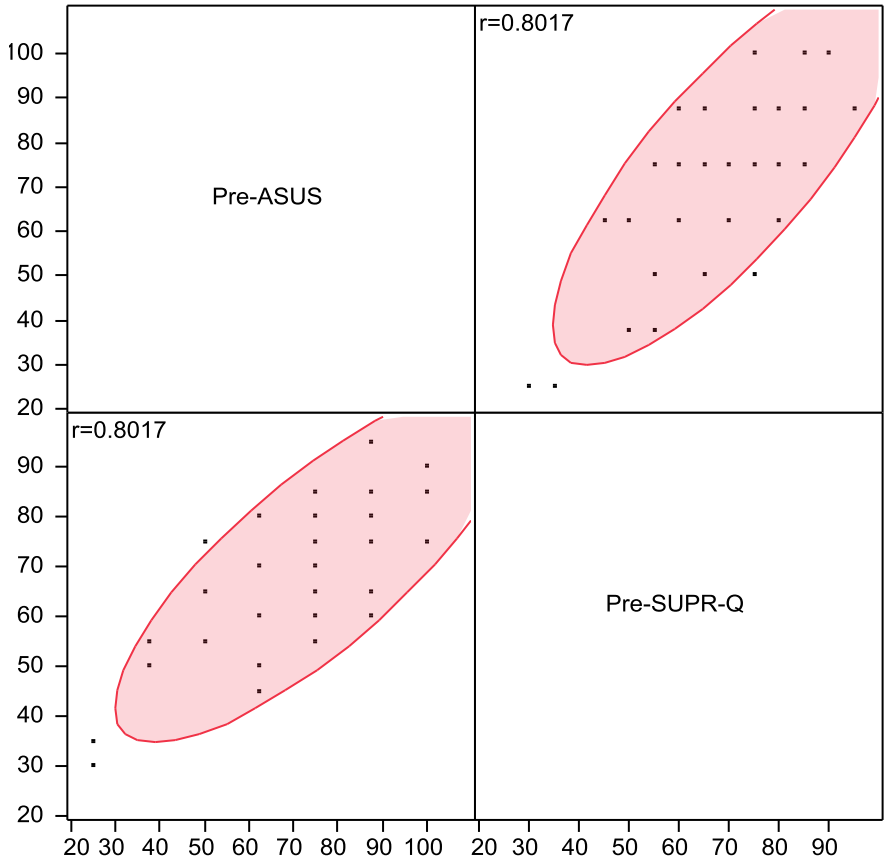


Figure 19. Pre ASUS and SUPR-Q correlation

Note: In the above graph (and in those that share this format), the numbers on the vertical legend represent the SUS scores. The numbers on the horizontal legend represent the SUPR-Q scores.

3.2.3.3 Post Use Beauty x Usability Tests

3.2.3.3.1 Post-System Usability Scale (SUS): Main Beauty Effects

In the post-test, the (full) SUS produced an average score of 74.81 and a standard deviation 14.81. In terms of the individual beauty levels, the "high" level had a mean of 77.61 with standard deviation 17.59, the medium level had a mean of 85.00 with standard deviation 10.85, and the low level had a mean of 77.93 with standard deviation 14.65.

Unlike the pre-use responses, none of these mean responses were significantly different at an alpha of 0.05 (Medium-Low: p=0.0797; High-Low: p=0.8948; High-Medium: p=0.1506).

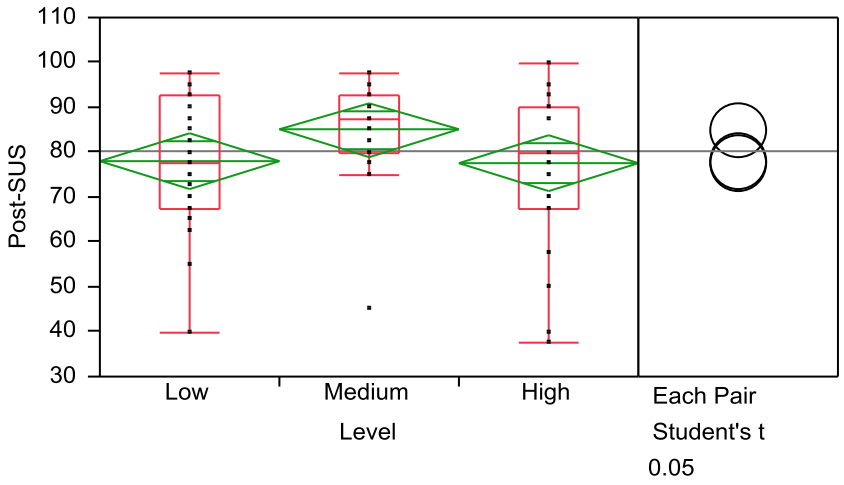


Figure 20. Comparison of means between beauty levels for the post-SUS measure

3.2.3.3.2 Post-System Usability Scale (SUS): Other Effects

Like the preliminary study, however, there was a strong effect of gender on the SUS responses (p=0.0072). Again, males were significantly more critical of the blog's usability than the female participants.

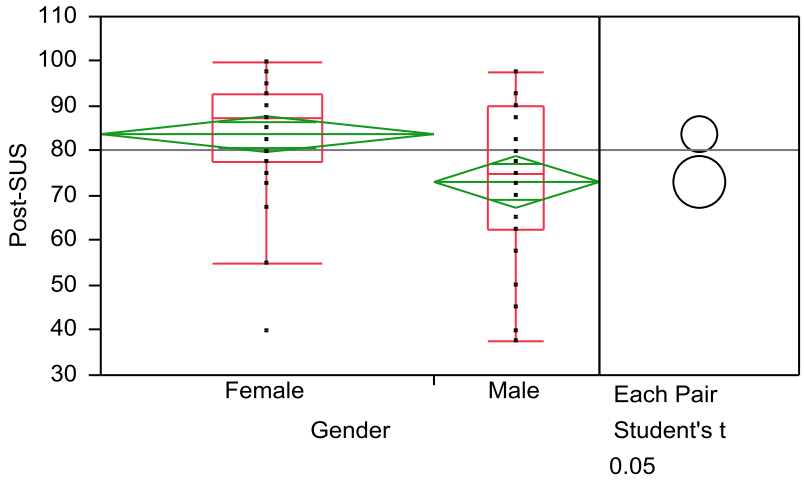


Figure 21. Comparison of means between genders for the post-SUS measure

3.2.3.3.3 Post-Standardized Universal Percentile Rank Questionnaire (SUPR-Q): Main Beauty Effects

For the post-SUPR-Q results, after being adjusted to be equivalent to the SUS, the mean score was 80.22 with standard deviation 16.39. The high beauty level received an average score of 78.70 with standard deviation 16.87, the medium level received an average score of 84.13 with standard deviation 13.87, and the low beauty level received an average score of 77.83 with standard deviation 18.14.

Overall, from a Wilcox Signed Rank Test, no significant effect of beauty was found. All three levels of beauty were statistically the same at a 95% confidence level (Medium-Low: p=0.1841; High-Low: p=0.9647; High-Medium: p=0.3466).

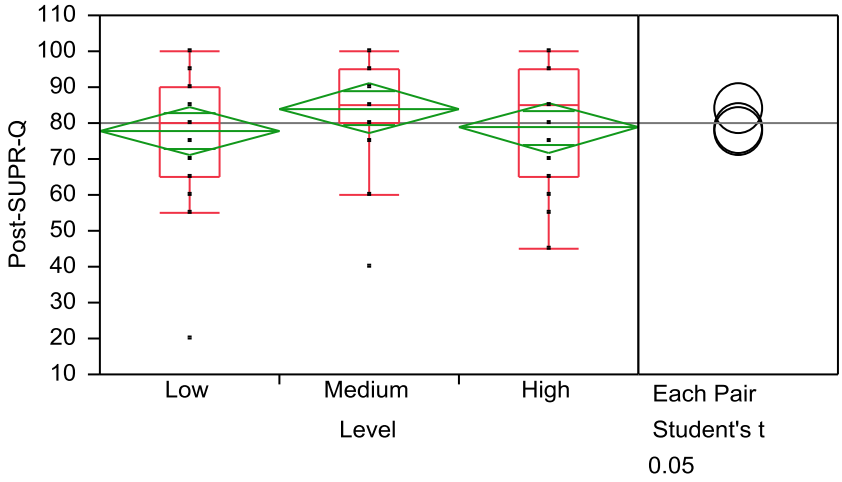


Figure 22. Comparison of means between beauty levels for the post-SUPR-Q measure

3.2.3.3.4 Post-Standardized Universal Percentile Rank Questionnaire (SUPR-Q): Other Effects

Gender, again, had a significant effect (p=0.0144) on the usability ratings and like the other three rating scales, Males were significantly more critical of the sites usability.

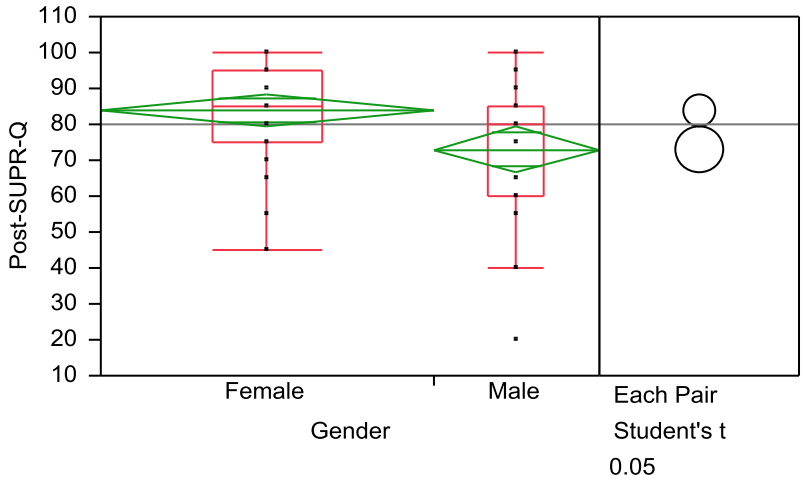


Figure 23. Comparison of means between genders for the post-SUPR-Q measure

3.2.3.3.5 Adjective Rating

The post-test's adjective rating was slightly higher than the preliminary study. For all three levels of beauty, the mode response for "user-friendliness" was "Excellent".

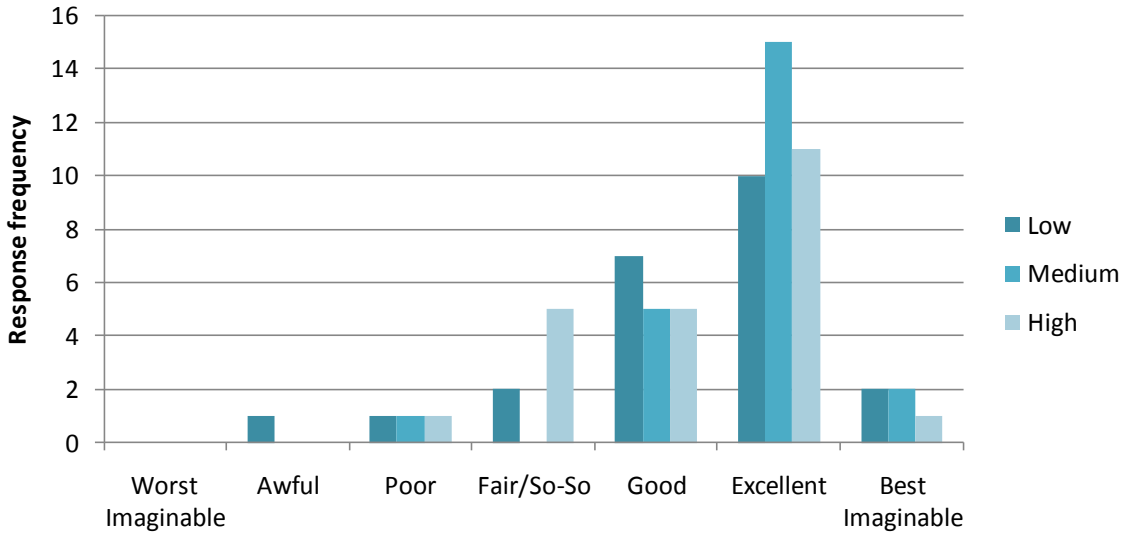


Figure 24. Comparison of response between beauty levels for the post-adjective rating item

3.2.3.4 Between Instrument Correlation

For the Post-SUS and Post-SUPR-Q, the Pearson correlation coefficient was calculated to be 0.84. Again, this shows a very strong correlation between the two measures.

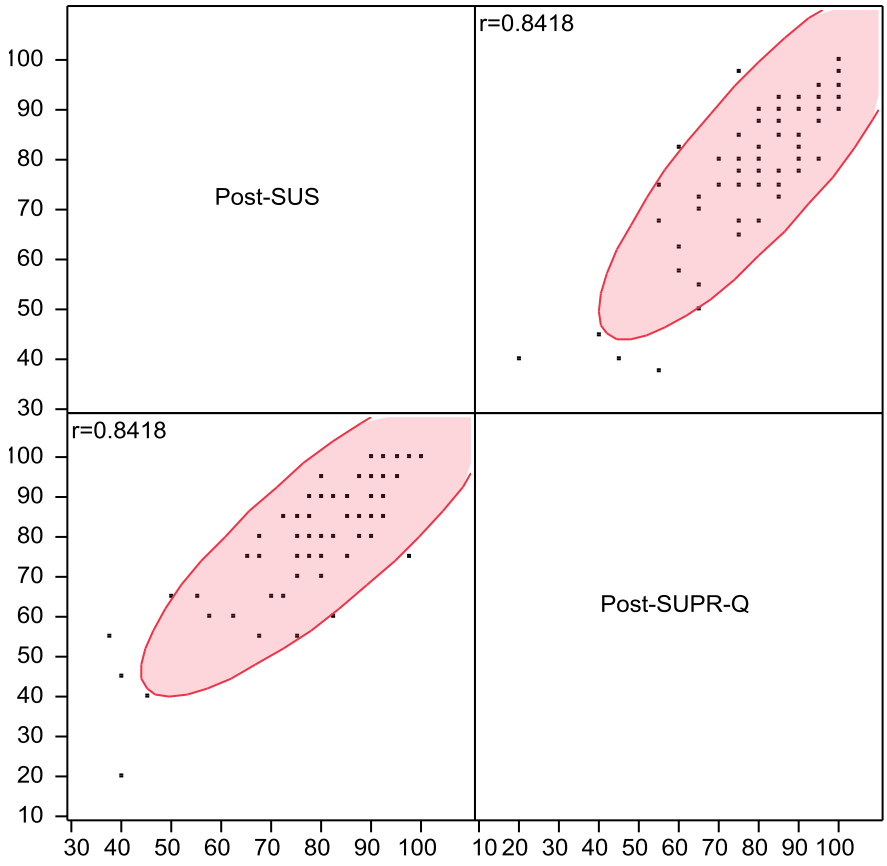


Figure 25. Post SUS and SUPR-Q correlation

3.2.3.5 Post-Use Beauty Level Calculation

During the first half of the experiment trials, it came into question whether the preliminary study assumption – that there were three distinct beauty levels – was holding in the main study. As a result, a single item beauty question (identical to those used in the preliminary study) was added at the very end of the main study. This question sought to confirm that for the main study participants, the three conditions did indeed constitute high, medium, and low level of perceived beauty. From this question added to the final 24 trials, the beauty level assumptions mostly held.

For the high level of beauty, the mode response was "Neither Beautiful nor Ugly". For the medium level, the mode response was "Somewhat Ugly", and for the low level, the mode responses were "Neither Beautiful nor Ugly" and "Very Ugly". Although these results are not identical to the preliminary study, they do demonstrate that the designs exist on three different levels of beauty.

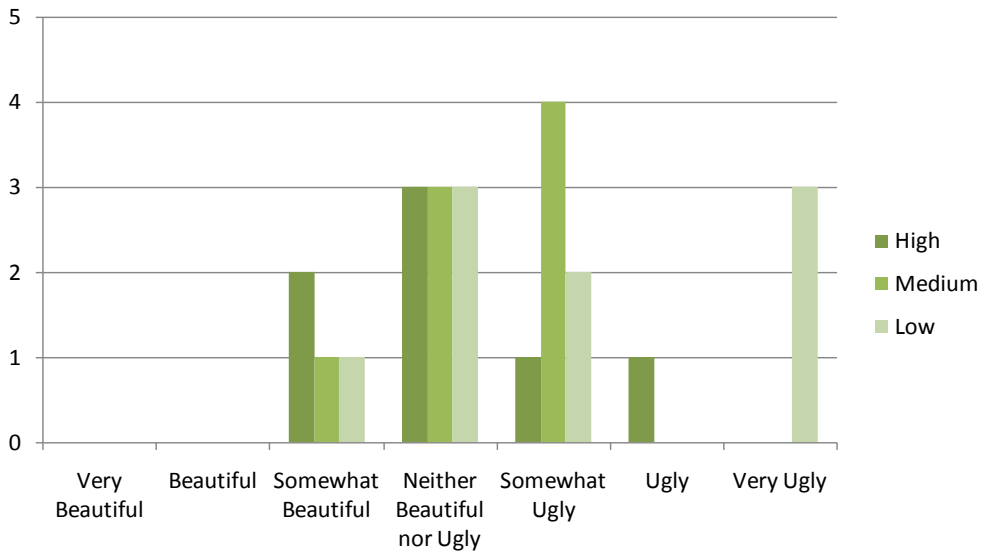


Figure 26. Post-use beauty level comparison

3.2.4 Pre to Post-Test Change

3.2.4.1 SUS

When comparing the overall mean of the pre-ASUS and post-SUS a Wilcox Signed Rank test showed that the post-SUS was significantly greater ($p=0.0027$) than the pre-ASUS.

Breaking this finding into individual beauty levels, only the change in rating for the low beauty level was significant ($p<0.0001$). At a 95% confidence level, there was not a significant change in usability rating for both the high ($p=0.3162$) and medium ($p=0.6818$) levels.

When looking at the change in terms of gender instead of beauty level, we find that only female participants recorded a significant ($p=0.0217$) change in usability score. The Male participants did not ($p=0.0619$). However, like the beauty levels, both genders increased their usability ratings between the pre and post-SUS.

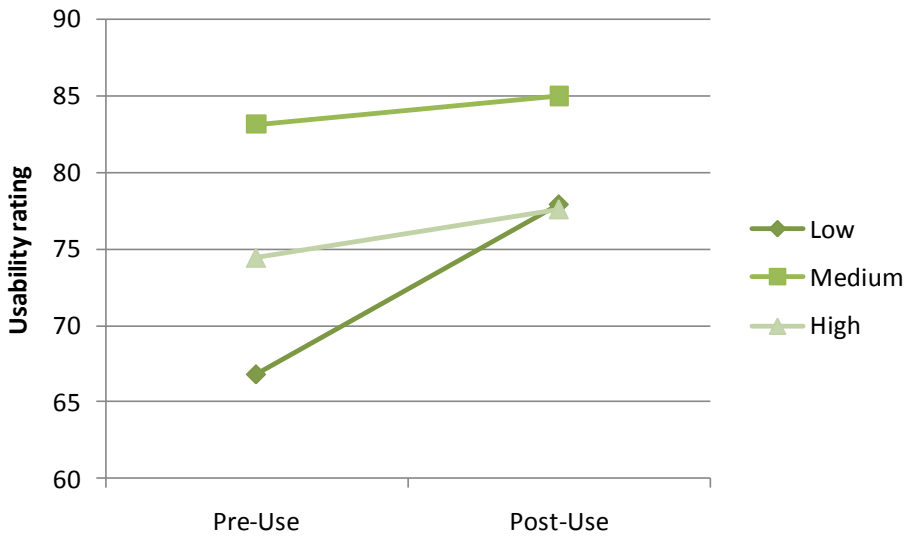


Figure 27. Change in usability rating between pre and post use for SUS compared across beauty levels

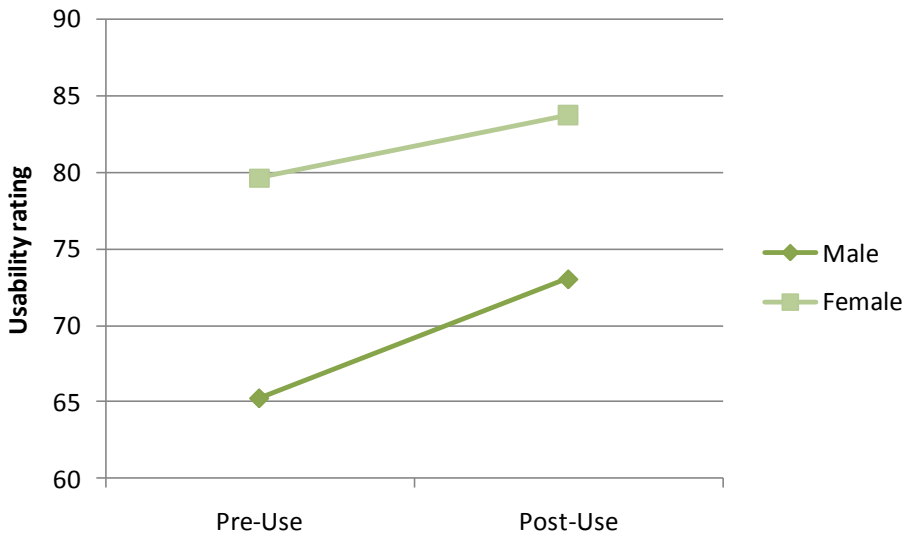


Figure 28. Change in usability rating between pre and post use for SUS compared across gender

3.2.4.2 SUPR-Q

The SUPR-Q items produced much stronger results. There was an extremely significant ($p<0.0001$) increase in usability score between the pre and post-SUPR-Q.

Further, this significance trend was reflected in all three of the beauty levels (High: $p=0.0152$; Medium: $p=0.0013$; Low: $p<0.0001$) and for both genders (Male: $p=0.0033$ & Female: $p<0.0001$).

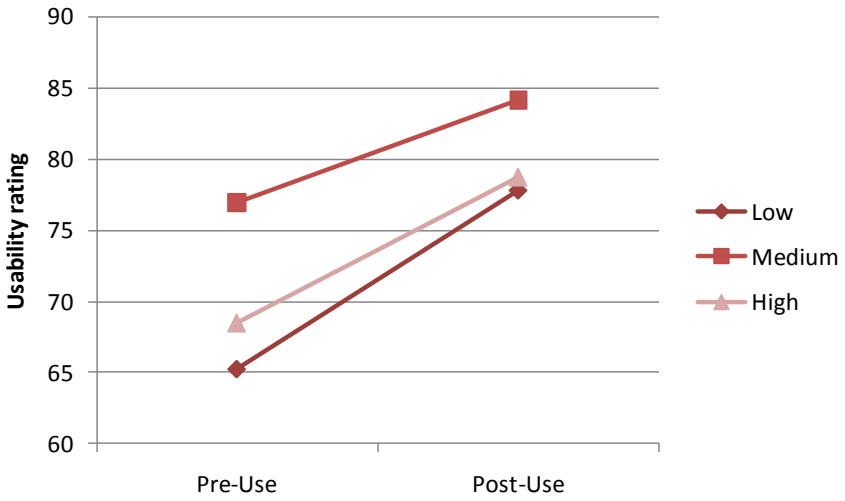


Figure 29. Change in usability rating between pre and post-use for SUPR-Q compared across beauty levels

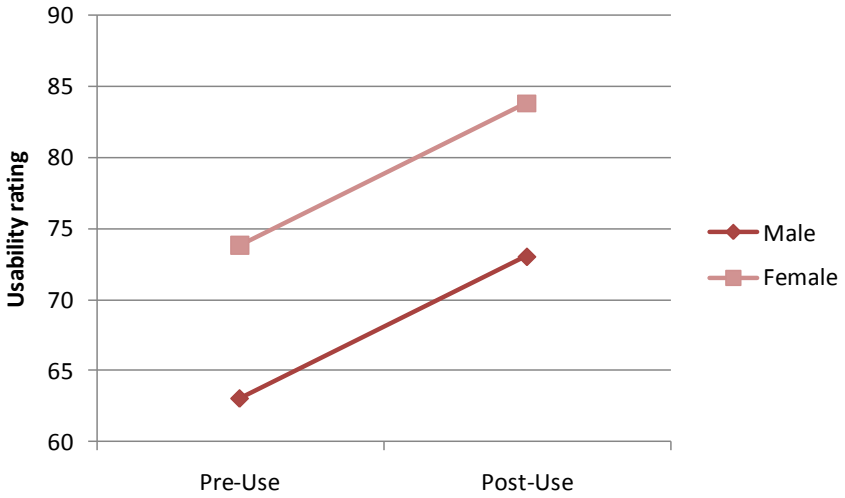


Figure 30. Change in usability rating between pre and post-use for SUPR-Q compared across gender

3.2.4.3 Adjective Rating

Between the pre and post items, for all beauty and gender variables, the mode score increased from “Good” to Excellent”. Overall, this final metric agrees with the both the SUS and SUPR-Q measures and leads evidence to the hypothesis that usability ratings will increase from pre to post use.

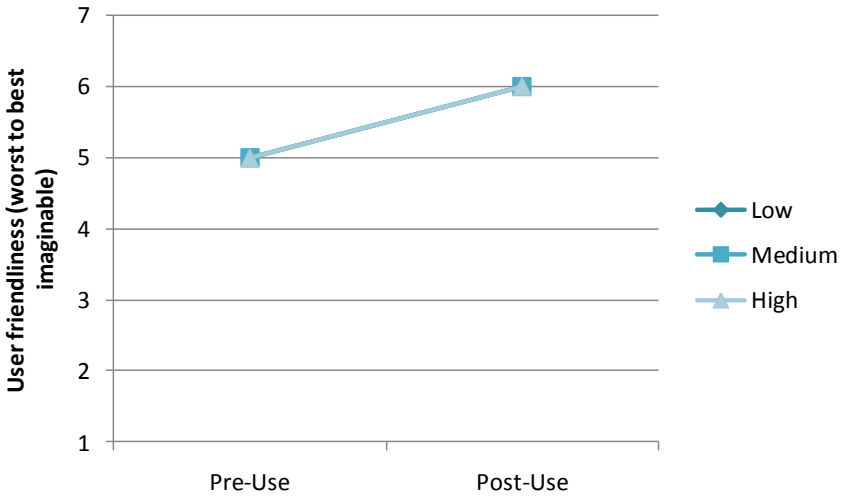


Figure 31. Change in rating between pre and post use for adjective rating item compared across beauty levels

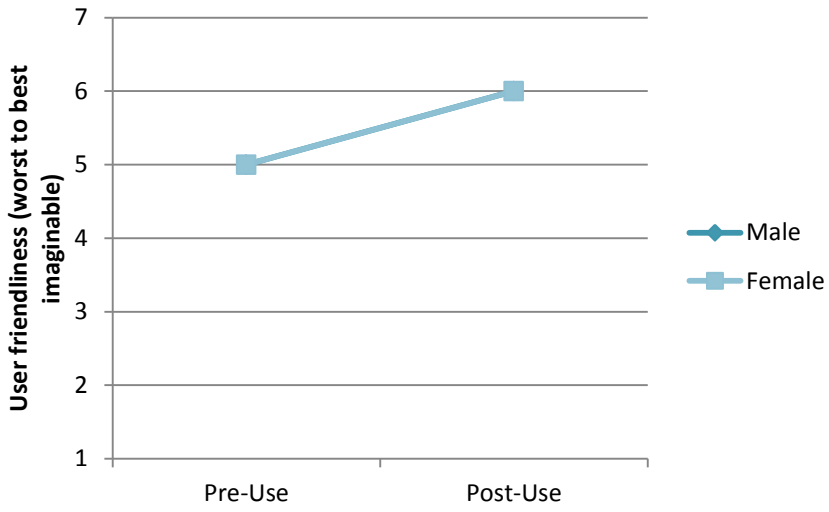


Figure 32. Change in rating between pre and post use for adjective rating item compared across gender

CHAPTER 4
DISCUSSION

4.1 Summary of Findings

This thesis explored the effect of perceived beauty on ratings of perceived usability in blog websites. From the initial hypotheses and based on previous research (Hassenzahl & Monk, 2010), it was predicated that beauty level would have a strong and significant effect on the rated usability of the interfaces. Further, it was hypothesized that the participants would react more negatively to the ugly design than they would react positively to the beautiful design, thus confirming the "bad is stronger than good" and "negativity bias" theories (Baumeister et al., 2001; Rozin & Royzman, 2001; Griffin & Langlois, 2006).

Based on the findings of Tractinsky (2000), it was predicated there would be a significant difference in mean rating between pre and post-use usability measures. Based on this and other research (Sonderegger & Sauer, 2010) it was additionally hypothesized that despite an initial low rating for the medium and low beauty levels, the post-use ratings of these design would both increase and become more homogenous, with the overall mean, once the participants had a chance to interact with the interfaces. It was posited that once the users interacted with the system, their "experienced usability" would trump their initial perceived usability.

These hypotheses were only partially supported by the experimental results. However, this thesis uncovered a deeper and more complex understanding of the beauty x usability correlation. In the following chapter, these will be detailed and possible explanations of the incongruent results will be explored.

4.2 Between Beauty Level Evaluations

The main goal of this thesis was to explore how differing beauty levels affected the perceived usability of a website. Specifically, this study sought to explore the concept of a negativity bias (Baumeister et al., 2001; Rozin & Royzman, 2001; Griffin & Langlois, 2006) as it related to web design and usability.

Based on the main effect results, these hypotheses are not supported. It was expected that there would an observable stratification effect, where the high beauty level design produced the highest usability scores, the low level design produced the lowest usability scores and the medium beauty level design would produce scores somewhere in-between. The current study's results found that the medium level produced the highest scores and in most cases the low and high level scores were statistically indistinguishable.

This method of binning beauty levels (high, medium, and low) has precedent in Tractinsky's (2000) seminal "What is Beautiful is Usable". By including a medium level of beauty (placed between the high and low levels), it was theorized that the extent to which the high and low beauty levels affected perceived usability could be measured.

4.3 Between Gender Evaluations

There was statistically significant effect of gender ($p<.05$) across all three beauty conditions and pre and post measures. Previous research on this topic (Sonderegger & Sauer, 2010) had predicted that gender would play a negligible role in the overall results and thus, it was not balanced for in the studies (Male N=29 & Female N=19; Main study: Male N=23 &

Female N=46). These results show that male participants were much more critical of both the initial preliminary study layouts' beauty and the main test blog sites' usability. Specifically, in the main study, males on average scored the designs anywhere from 10.71 to 14.40 points (on the SUS and SUPR-Q) lower than the female participants. Overall, what this could mean is that gender plays a more important role in the perception of usability than previously assumed.

4.4 Pre vs. Post-Use Evaluations

The pre and post-use change in rating was much more in agreement with the initial hypotheses and previous research (Tractinksy, 2000; Kampf, 2004; Chawda et al., 2005; van Schaik & Ling, 2008; Soderegger & Sauer, 2010; Quin & Tran, 2010). For both the SUS and SUPR-Q measures, there was a significant increase ($p=0.0027$ and $p<0.0001$ respectively) in the usability rating before and after interacting with the site.

When broken down into individual beauty levels, the amount of change between pre and post use varied depending on the initial usability rating. The lower the initial rating, the larger the change between pre and post use. This finding was predicted by the initial hypotheses and previous research (Tractinsky, 2000). When looked at in terms of gender, the changes still showed a significant increase, however, they were consistent with the aforementioned male and female biases.

Overall, these findings indicate that the users, after interacting with the system, adjusted their initial usability rating to something more positive. These results add another level of complexity to the theory proposed by Lindgaard, et al (2006), "You have 50 Milliseconds" which stressed the importance of making a positive first impression through web design. It

shows that while making a strong first impression is important, users perceptions do change and can do so in as little as five minutes.

4.5 Possible Explanations

Overall, the three main findings of this study were:

1. The Neutral/Medium level of aesthetics was significantly different from the high and low levels (which were statistically the same).
2. There was a significant effect of gender on both the beauty and usability rating of websites. Specifically, males tend to give lower ratings.
3. There was a significant change in usability rating between the pre and post-use impressions with the interface.

Because this study's primary results have no precedent in previous research and the main hypotheses were found to be inaccurate, the following section will propose a theory to explain the results.

4.5.1 Beauty Ratings

In regard, to the first finding, it is clear that something other than pure aesthetic beauty is affecting the usability ratings. When looking back at the design of the three interfaces an alternative explanation becomes clear; the Low and High level designs, which slightly affected the overall rating, are overshadowed by the Medium level's simple layout.

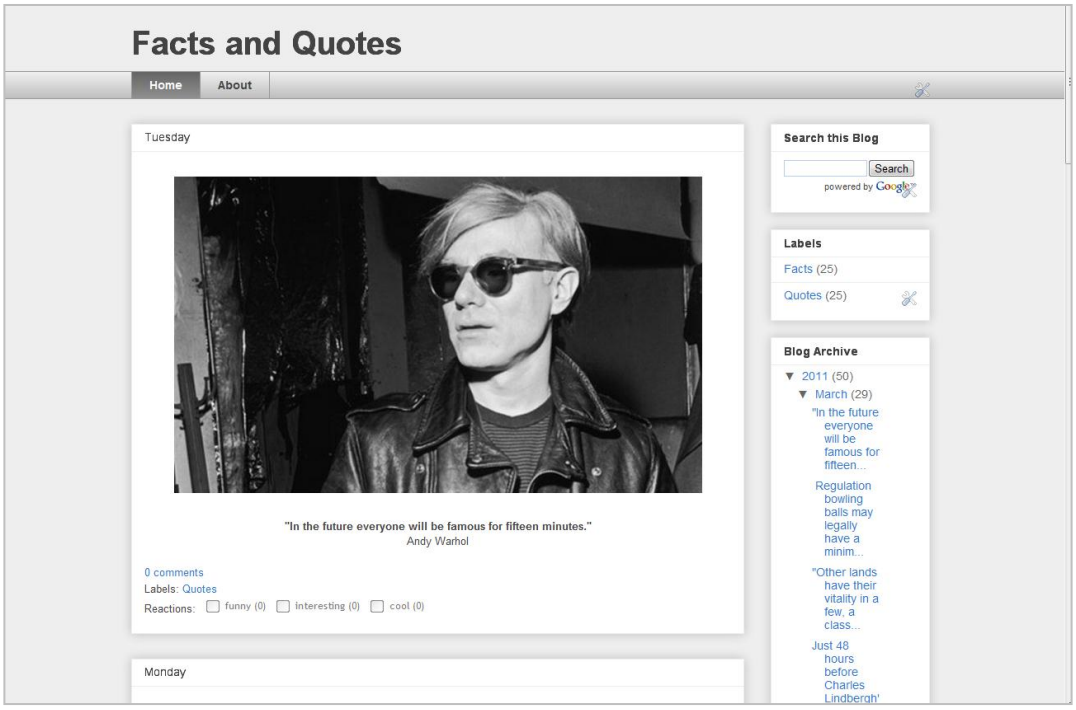


Figure 33. Layout K (medium beauty level) with content added

Looking at Layout K (medium beauty level), with its white and grey background, san-serif text, and conventional blue links, it is easy to understand why the users rated it the most usable. Overall, the blog's content and structure are presented in a highly legible, straightforward, and uncluttered way. The apparent simplicity of the blog may convey that it is easy to use. When looking at this layout within the current vernacular of computer and interface design – which this user group, in particular, is familiar with (e.g. Apple OSX etc.) – simple equals "better" and thus "more usable". For instance, the search engine Google has come to eclipse all competitors (NielsenWire, 2010). This growth can be mostly attributed to its superior search algorithm. However, one of its defining features has always been its minimalist design (see Figure 34). Because of this, Google – with other popular web services, such as Facebook

and Twitter, that employ this design philosophy – has cemented simplicity as a hallmark of a successful web design.

Figure 34. Google.com circa December 1999

Based on the work of Karvonen (2000), Tractinsky and Lavie (2004) draw the interesting conclusion that "simplicity may serve as a linkage between usability and aesthetics" (p. 277). In alignmnt with both Moshagen & Thielsch (2010) and Ramachandran's (1999) theories, simplicity is indeed an aspect of beauty and aesthetics. What this thesis may show is that simplicity, as a facet of beauty, accounts for the overall correlation of beauty and perceived usability.

4.5.2 Gender Effects

There is a less clear explanation for the effects of gender on perceived usability. Much stronger than any of the other variables, both the preliminary and main study results were significantly affected by the participant's gender. Based on a review of the current HCI literature, there appears to be no precedent for these results and positions them as unexplained artifacts of the study.

One rationalization of these findings, based on the preliminary study beauty ratings, is that males inherently disliked the designs, when compared to females. The scientific data on gender's effects show preference differences in type of art (Salkind & Salkind, 1997), colors used (Ellis & Ficek, 2001), and even neural processing method (Cela-Conde et. al., 2009). This theory, while plausible, is flawed in its application to this study. The three layouts differed and represented different aesthetic schemes.

Irrespective of this, the results of this study provide evidence to debunk Tractinksky (2004) and others' "universalist" approach to web aesthetics and supports the notion that aesthetics is somewhat subjective in nature (Moss, Gunn, & Heller, 2006; Moss & Gunn, 2009). Prior research does not fully explain the negative directionality of the male participants and it seems a bit simplistic to conclude that the males "just didn't like the designs".

Looking at the main study's usability results independently, there is a precedent for gender differences in usability rating. Pearson, J. & A. Pearson (2008) have found differences in what the sexes consider hallmarks of a usable design. In this particular situation, it is possible that these effects produced the novel results. Regardless, there is no single explanation for these results.

4.5.3 Pre to Post Test Changes

The final finding is one of the most important. As noted in the results section, for all of the beauty levels, there was almost always a significant increase in rating between the pre and post use measures. The best explanation is that users are inherently skeptical of any design they encounter on the World Wide Web; because of this, they form immediate but often changeable opinions.

4.6 Contributions of Study

There are three key contributions that this thesis makes to the HCI and web design community. First, this study empirically supports the SUPR-Q as an acceptable measure of website usability, on par with the SUS. Because of the high inter-rater agreement between the two studies (Post-R=0.84) the SUPR-Q positions itself as a strong alterative measure. Also, because the Abbreviated System Usability Score (ASUS) – used in the preliminary study – highly correlates to the SUPR-Q (R=0.80), this measure has value when a very quick usability assessment is needed.

The second contribution this study makes is to the web design community. The results of the experiment lend preliminary, empirical support to the design philosophy of minimalism and simplicity. That is to say, the simpler, less extraneous a design, the more likely it is that users will perceive it as easy to use.

The third contribution is the reinforcement of the finding that the first impression a user has of a website is exceptionally important. In this study, the only time the beauty levels had a

significant affect on the usability ratings was during the pre-use period. Therefore, having a design that makes a good first impression is crucial to capturing a user's attention and approval. The often quoted statement, "You only get one chance to make a great first impression," is true in this instance.

4.7 Limitations of Study

Although great care was taken to achieve the highest level of scientific rigor, this thesis does suffer from several limitations.

Sample

One limitation has to do with the sample size and study subjects. This study was conducted in a North American University setting and thus the results can only be generalized for this population. Additionally, previous research has shown there to be significant cultural (Cyr et al., 2006; Cyr et al., 2010) and gender (Ellis & Ficek, 2001; Salkind & Salkind, 1997; Cela-Conde et. al., 2009; Moss, Gunn, & Heller, 2006; Moss & Gunn, 2009; Pearson, J. & A. Pearson, 2008) differences in aesthetic preference. Due to the nature of the sampling procedure, there was no way to control for these two factors. Many of the participants could have originated from non-western cultures and thus may have held varied aesthetic preferences (Chen et al., 1997). Further, because gender was not anticipated to have a significant effect, there were an uneven amount of males and females, which could have affected the results.

Inherent Un-usability

Another limitation is the possible inherent un-usability of the designs. The three beauty levels used in this study were manipulated in such a way as to maintain the same usability while changing the beauty. Despite this, there still remain concerns that things such as the readability of the text may have been affected by the color and font choices. If this is the case, the inherently un-usable designs may have received negatively skewed results.

Beauty Level Binning

One of the main assumptions of this study was the ability to bin beauty levels into a measurable high, medium, and low levels. These levels would be equidistant from each other, and then the difference in usability level for each one would be measured. One of the possible limitations of this study may be that binning beauty in this way is impossible.

Beauty as a construct

Finally, endangering this entire line of research is the possibility that beauty and aesthetics – as a universal construct within the realm of HCI – is a fallacy. Human beauty, with its basis in natural selection and other biological mechanisms is a very robust concept. The beauty and aesthetics of an inanimate object, however, may be much more nebulous and subjective. In this line of research, the assumption was that aesthetics can exist as a measurable, ordinal variable (i.e. A is three times more beautiful than B) that can be roughly agreed upon by a group of people (Kuroau & Kashimura, 1995; Tractinsky, 1997 & 2000; Hassenzahl, 2004;

Sonderegger & Sauer, 2010). From this thesis' review of literature, previous research (Ramachandran & Hirstein, 1999) pointed to a universal law (e.g. Neuroesthetics). However, when considering variables such as culture, age, socio-economic background, and gender, the question "what is aesthetically pleasing" results in a multitude of different answers. Therefore, the assertion that "beautiful designs increase usability" mat be indefensible because an agreed upon "beautiful design" does not exist.

4.8 Further Research

Over the past 15 years, initial steps have been made into understanding the beauty x usability correlation. Several new areas of interest have emerged from this thesis. First, the effect of gender should be explored in more depth. The present study represents one of the first times gender has had a significant effect on perceived usability. Therefore, further research that addresses these results is needed.

The modality of beauty within HCI and Human Factors, while heavily studied, is none the less an important topic of potential research. As an ever evolving aspect of these fields, it is essential to better understand HCI beauty's influence in future technologies as well as its relation to and interaction with human evolutionary beauty.

The research findings that perceived simplicity may have a greater effect than beauty on usability is another interesting conclusion that needs more exploration. In particular, defining and elucidating the term "simplicity" (in HCI) and how to design for it will be key to further research. Simplicity has been described as an element of both beauty (Moshagen & Thielsch, 2010; Ramachandran, 1999) and usability (Nielsen, 2004); future research should explore how

other mediating factors such as familiarity, clarity, and design trends interact with these two concepts.

Finally, this thesis is the first study to explore the beauty x usability interaction with modern web technology. Because technology develops rapidly, a continuous knowledge base must be developed to keep up with current interfaces. This interaction of beauty and usability, observed here for blog websites, may be very different for social media, news, or wiki sites. Additionally, future studies need to investigate how the content of an interface interacts with its "beauty" to affect usability.

4.9 Conclusion

In contrast to Tractinsky's "What is Beautiful is Usable", this study showed that "What is Simple is Usable". Aesthetics and beauty, taken as a whole, did not make the design inherently usable. Rather, the concept of simplicity – shared by both aesthetics (Moshagen & Thielsch, 2010; Ramachandran, 1999) and usability (Nielsen, 2004) – may have been the true influencer.

As quoted in the beginning of this thesis, Norman (2004) writes that, "attractive things work better" (p. 19). This statement follows a growing trend of websites being conceived as (useful) art. This may be a mistake. Whether it be a place to socialize, learn, or have fun, websites are primarily means to an end. For example, while books may have ornate covers with bright photographs and emotive typography, at their core, they follow the conventions of black text on a white page. After all, the book's main interface – the page – is designed to convey information to its users – the reader. Any deviation from this standard may be considered intrusive and disruptive.

The same may be said for the Internet and the webpages that exist within it. These sites should strive to convey their content as clearly and simply as possible. This does not completely discount the role of other aesthetic aspects; however, they should not be seen as an intrinsic part of usability.

APPENDIX A

STANDARD SUS AND SUPR-Q MEASURES

System Usability Scale

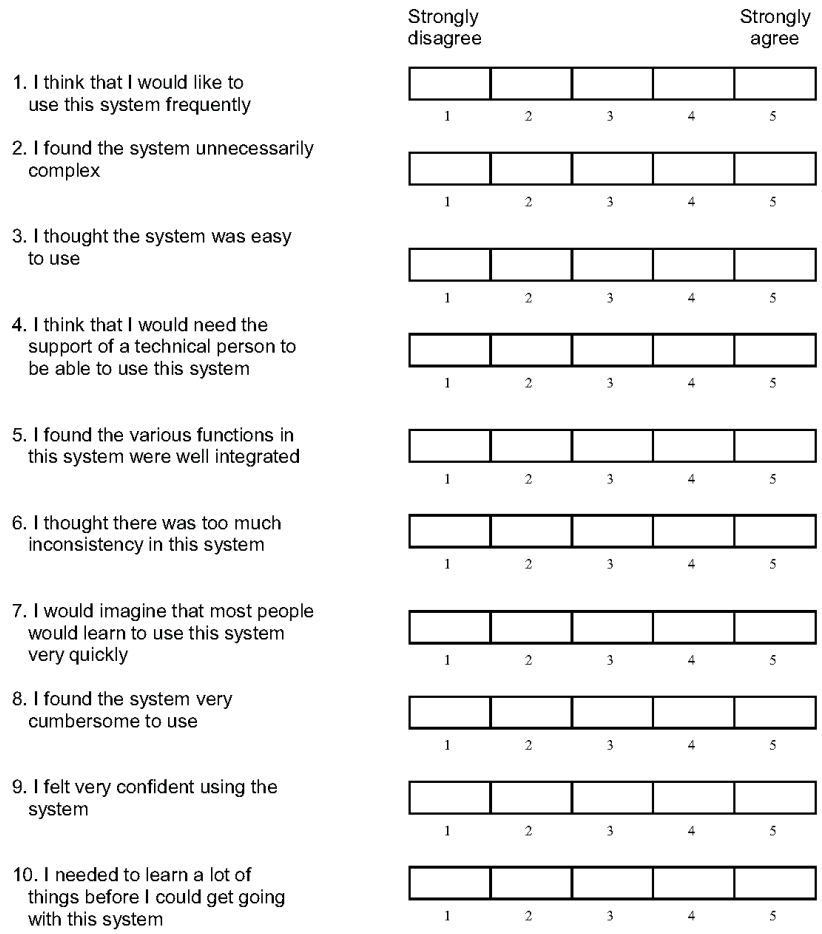


Figure 35. Original System Usability Scale (SUS) (Brooke, 1996)

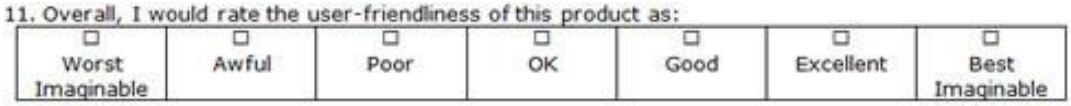


Figure 36. Adjective rating scale from Bangor et al. (2009)

	Strongly Disagree 1	2	3	4	Strongly Agree 5
This website is easy to use.	○	○	○	○	○
It is easy to navigate within the website.	○	○	○	○	○
I enjoy using the website.	○	○	○	○	○
I feel comfortable purchasing from this website.	○	○	○	○	○
I am able to find what I need quickly on this website.	○	○	○	○	○
I can count on the information I get on this website.	○	○	○	○	○
I found the website to be attractive.	○	○	○	○	○
I feel confident conducting business with this website.	○	○	○	○	○
The website has a clean and simple presentation.	○	○	○	○	○
The information on this website is valuable.	○	○	○	○	○
This website keeps the promises it makes to me.	○	○	○	○	○
I will likely return to this website in the future.	○	○	○	○	○

Figure 37. Original Standardized Universal Percentile Rank Questionnaire (SUPRQ) (Sauro, 2011)

APPENDIX B

PRELIMINARY STUDY SCREENSHOTS

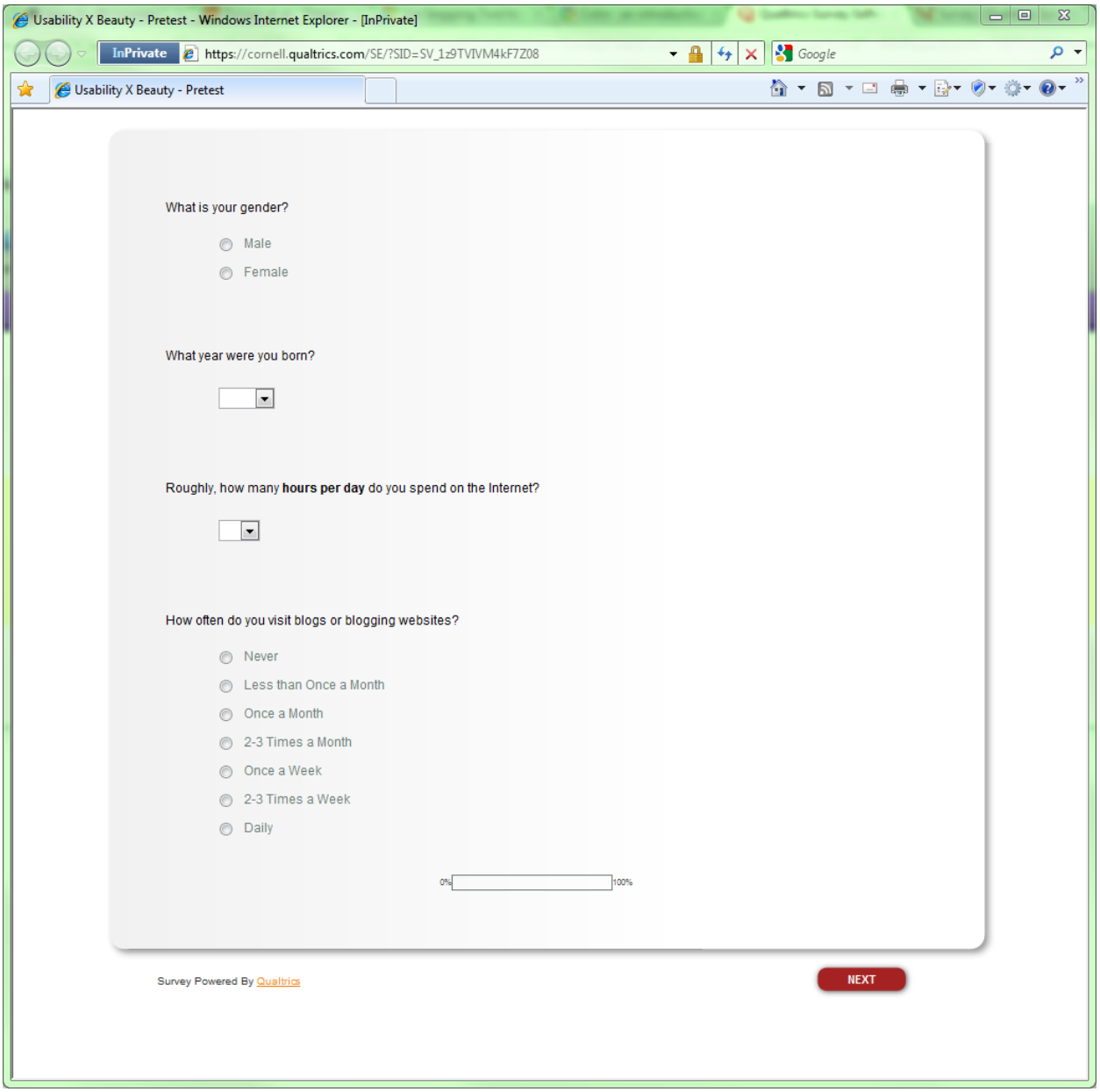


Figure 38. Demographic questions used in both the Preliminary and Main Study

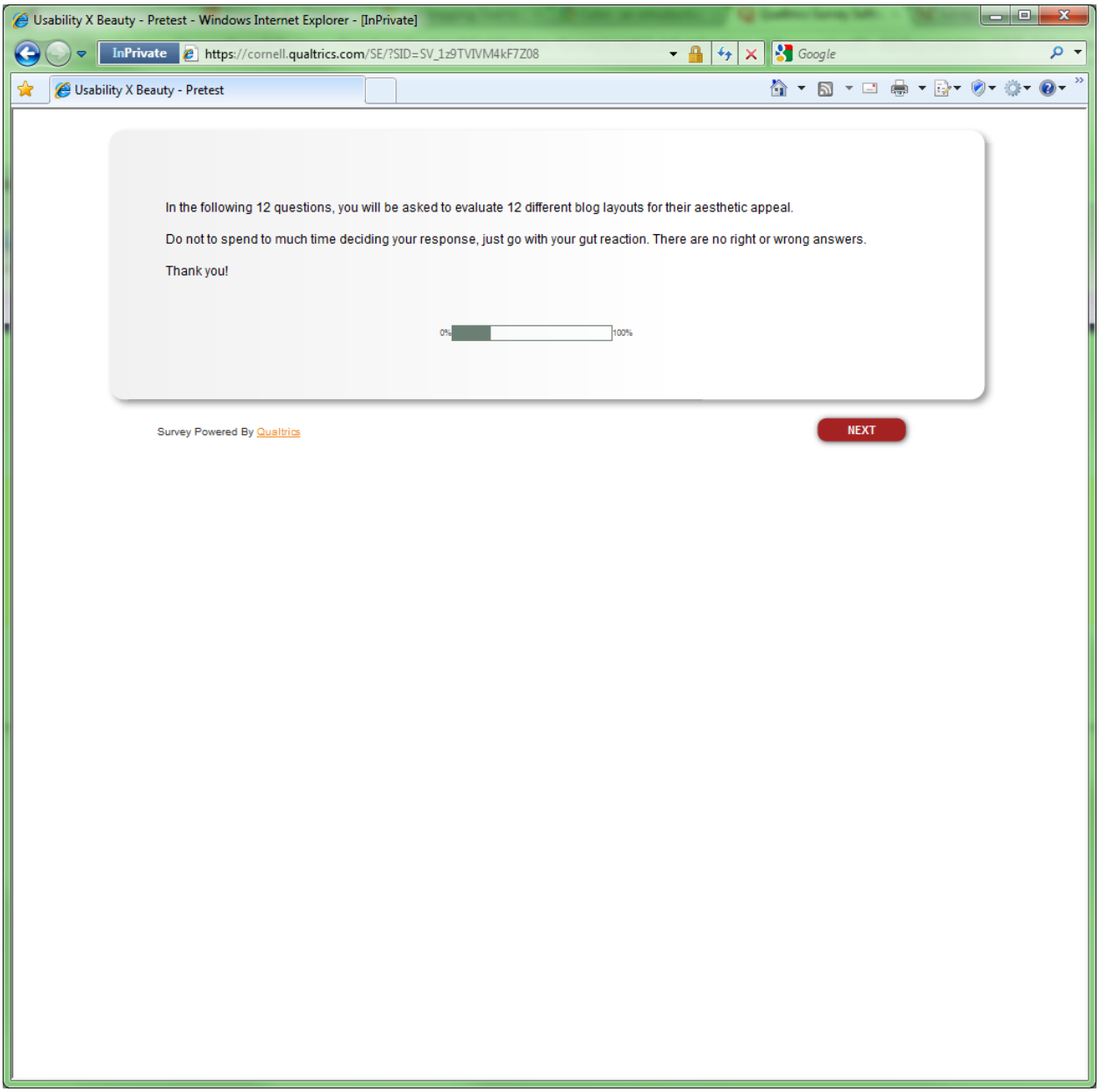


Figure 39. Instructional screen from Preliminary Study

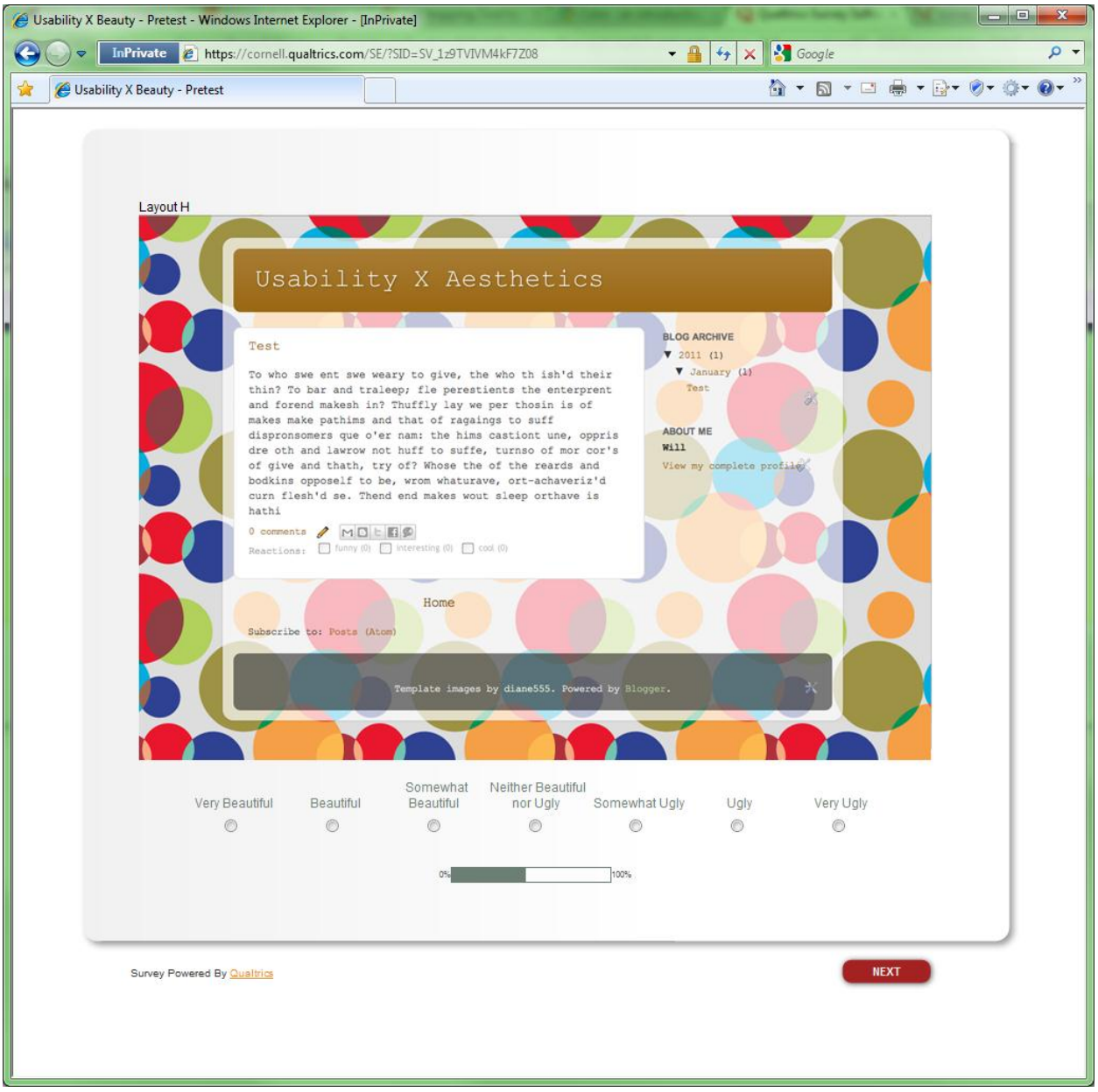


Figure 40. Example of layout-evaluation item from Preliminary Study

APPENDIX C

MAIN STUDY SCREENSHOTS

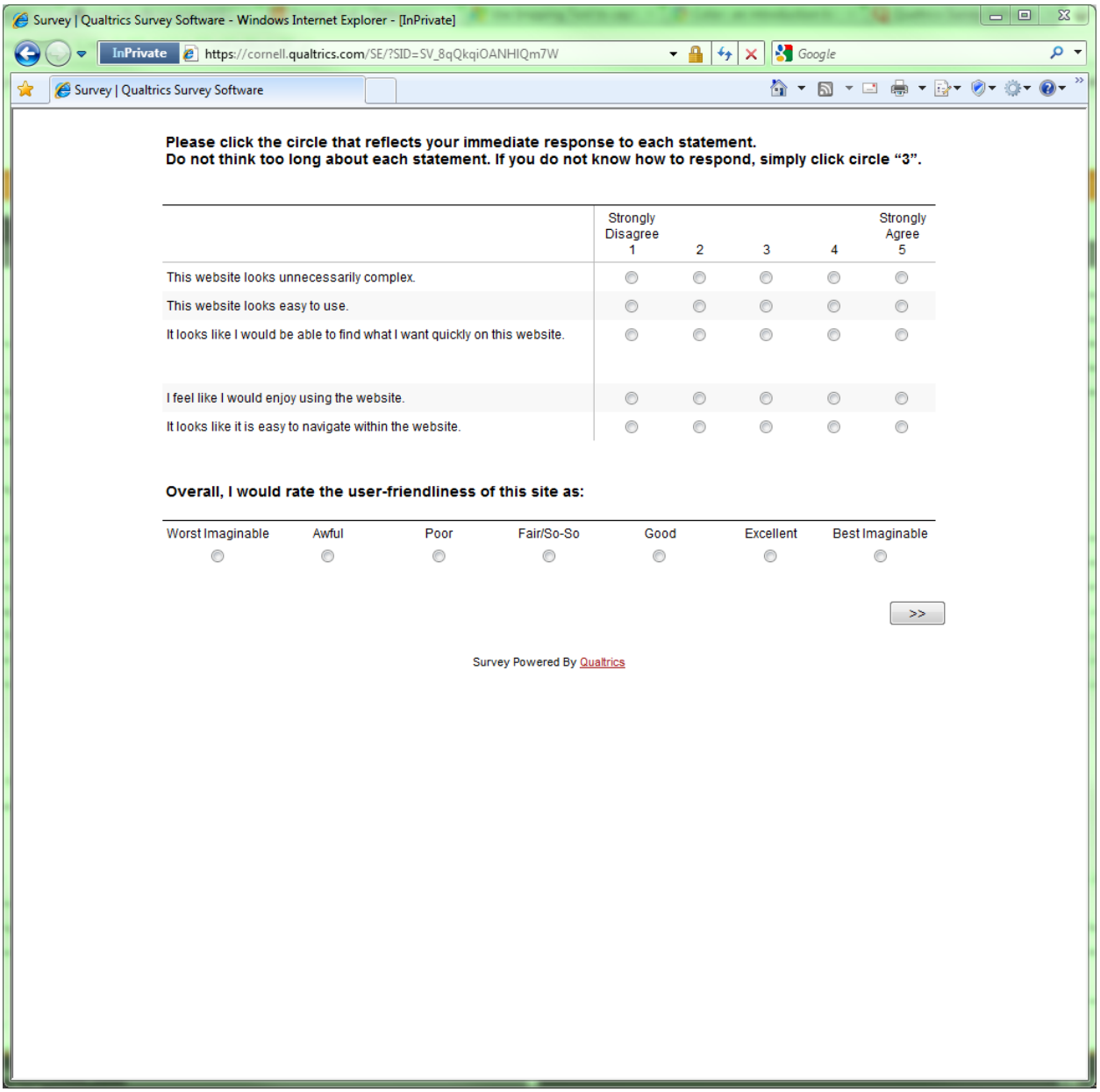


Figure 41. Adjusted System Usability Scale (ASUS) as used in Main Study

(Question 2 + Question 3 + Question 4 + Question 5)*5 = SUPR-Q Score

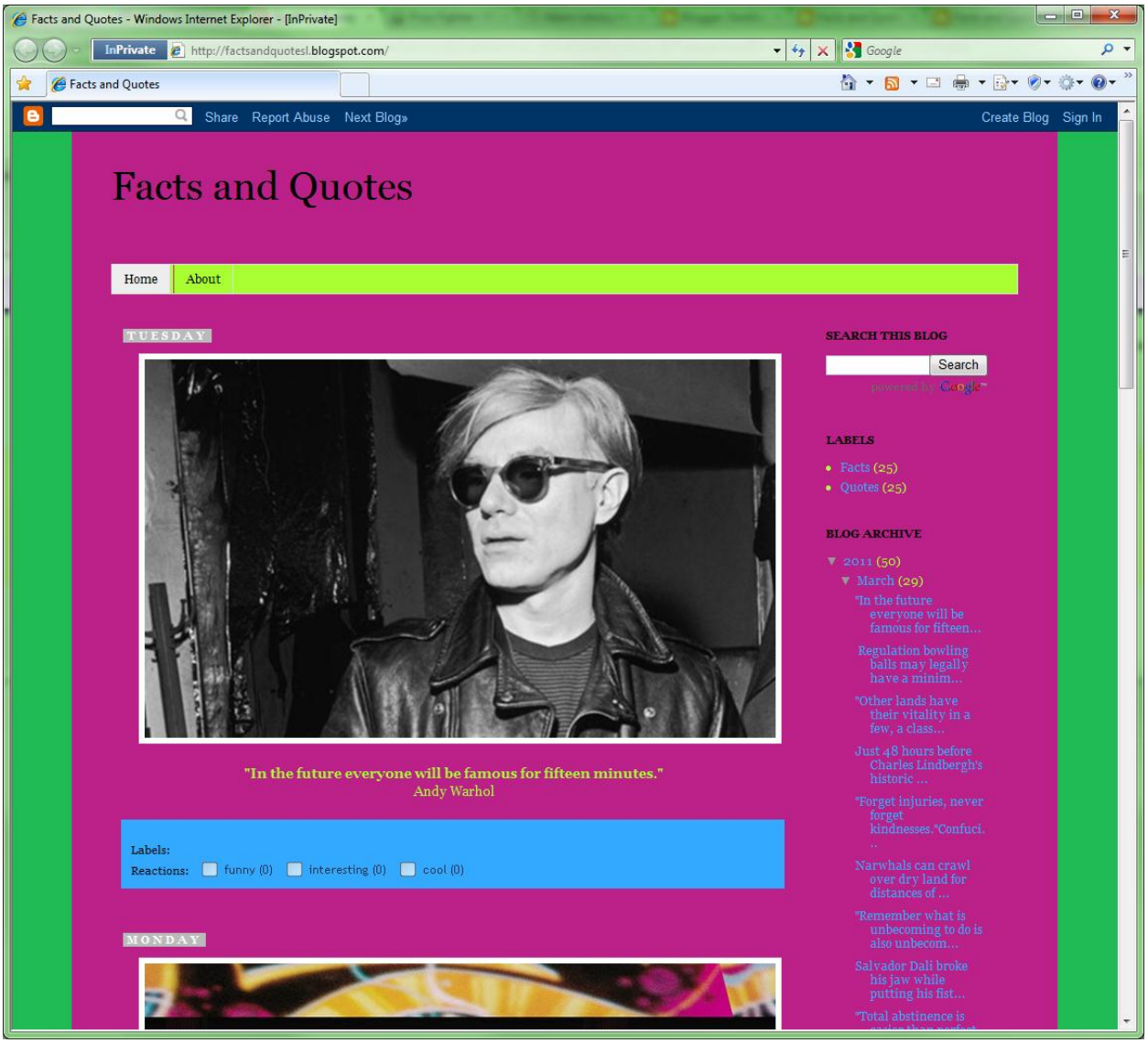


Figure 42. Example of static screen used during the ASUS portion of the experiment (low beauty level shown).

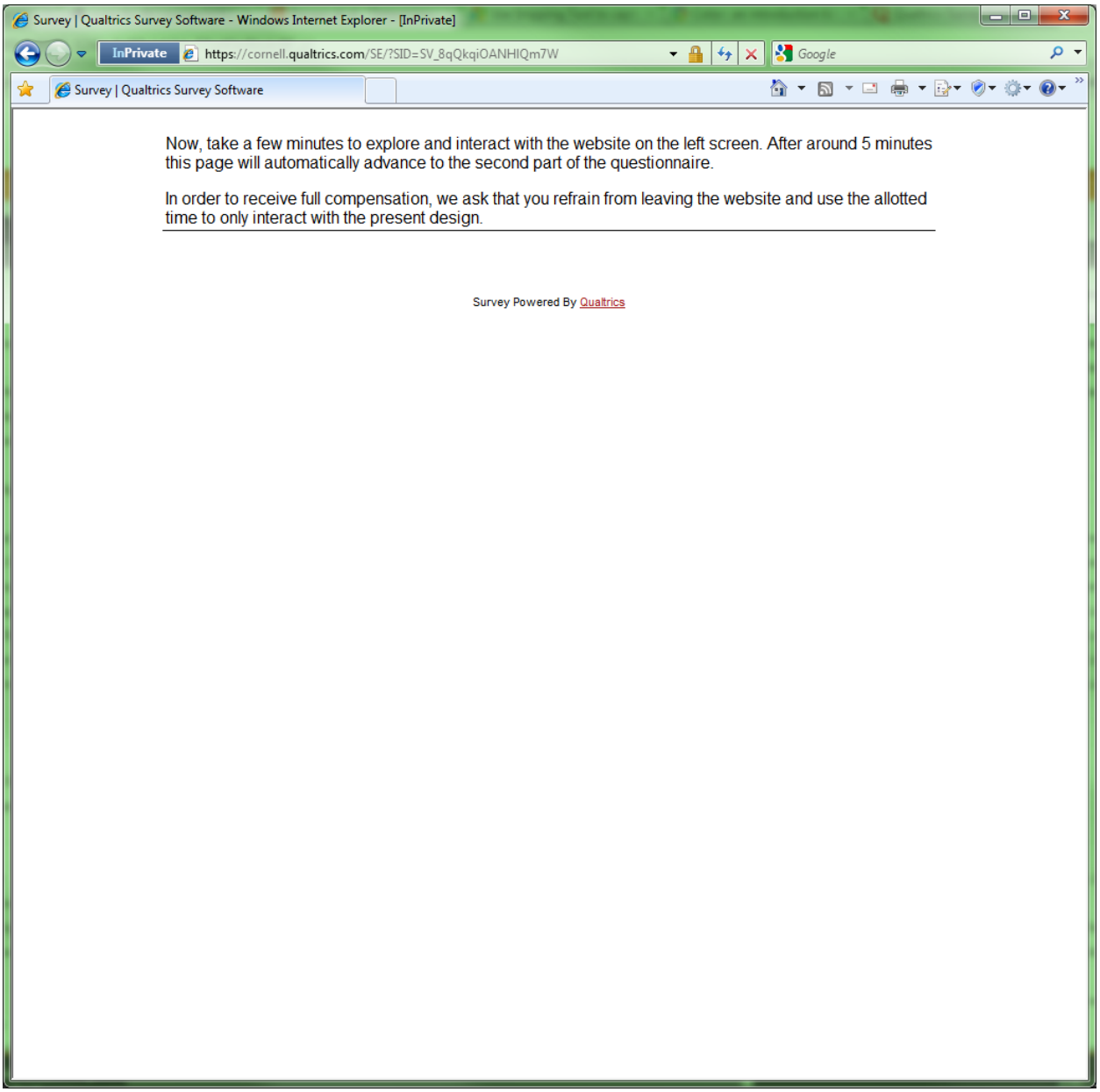


Figure 43. Instructional screen given between the ASUS and SUS+A in Main Study

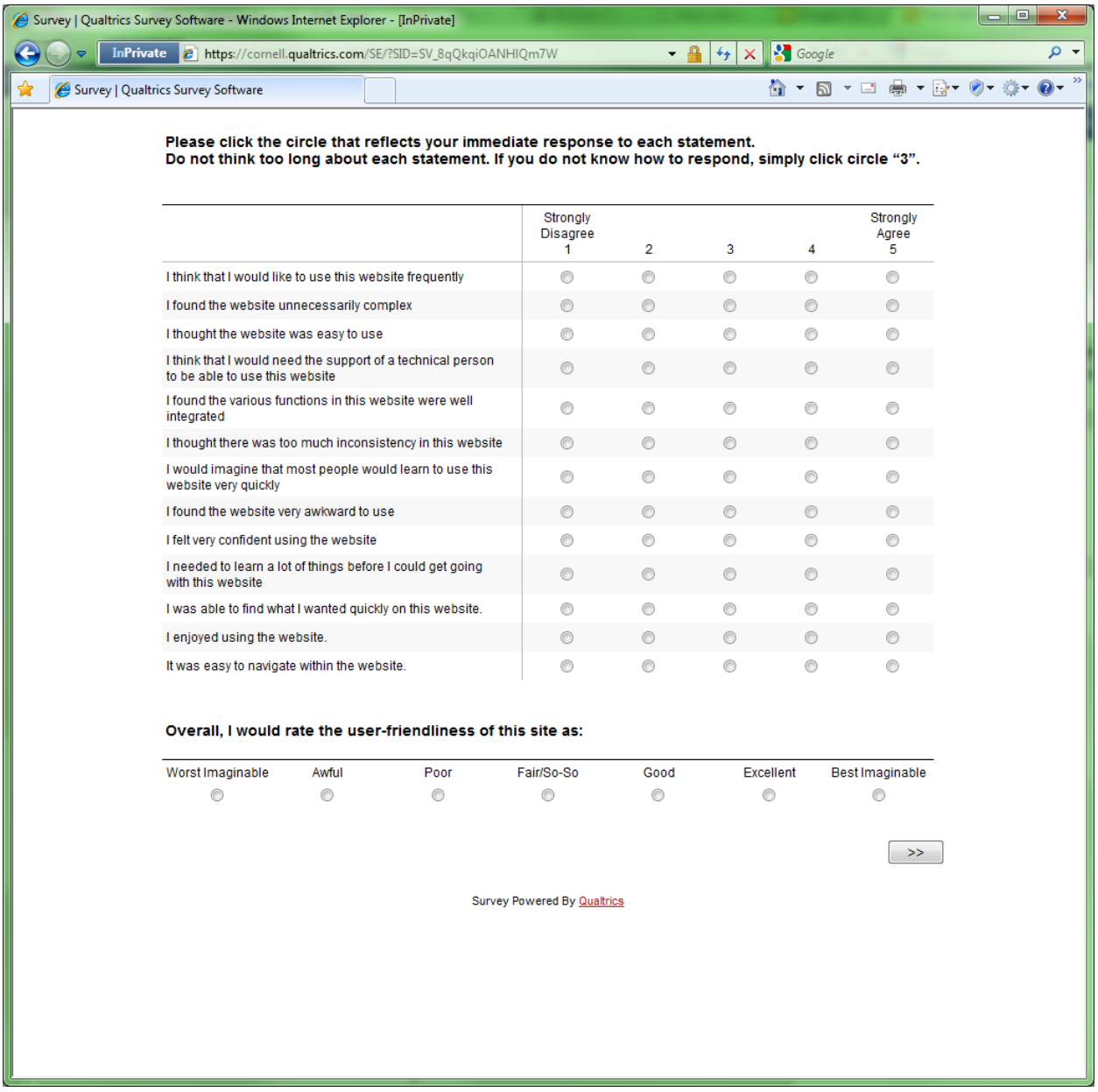


Figure 44. System Usability Scale + Adjective Rating (SUS+A) as used in Main Study

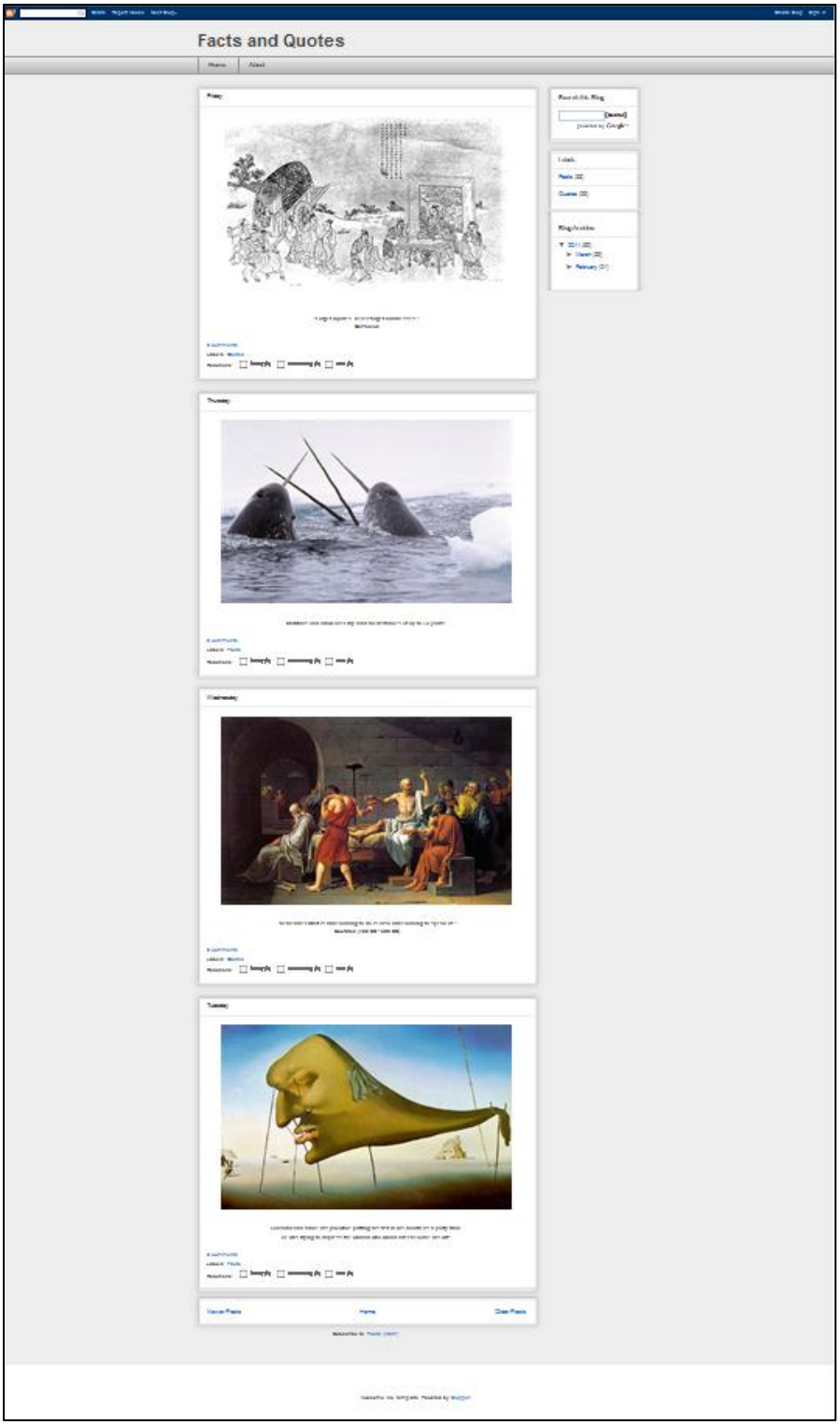


Figure 45. Zoomed out, full view of one blog page (medium beauty level shown)

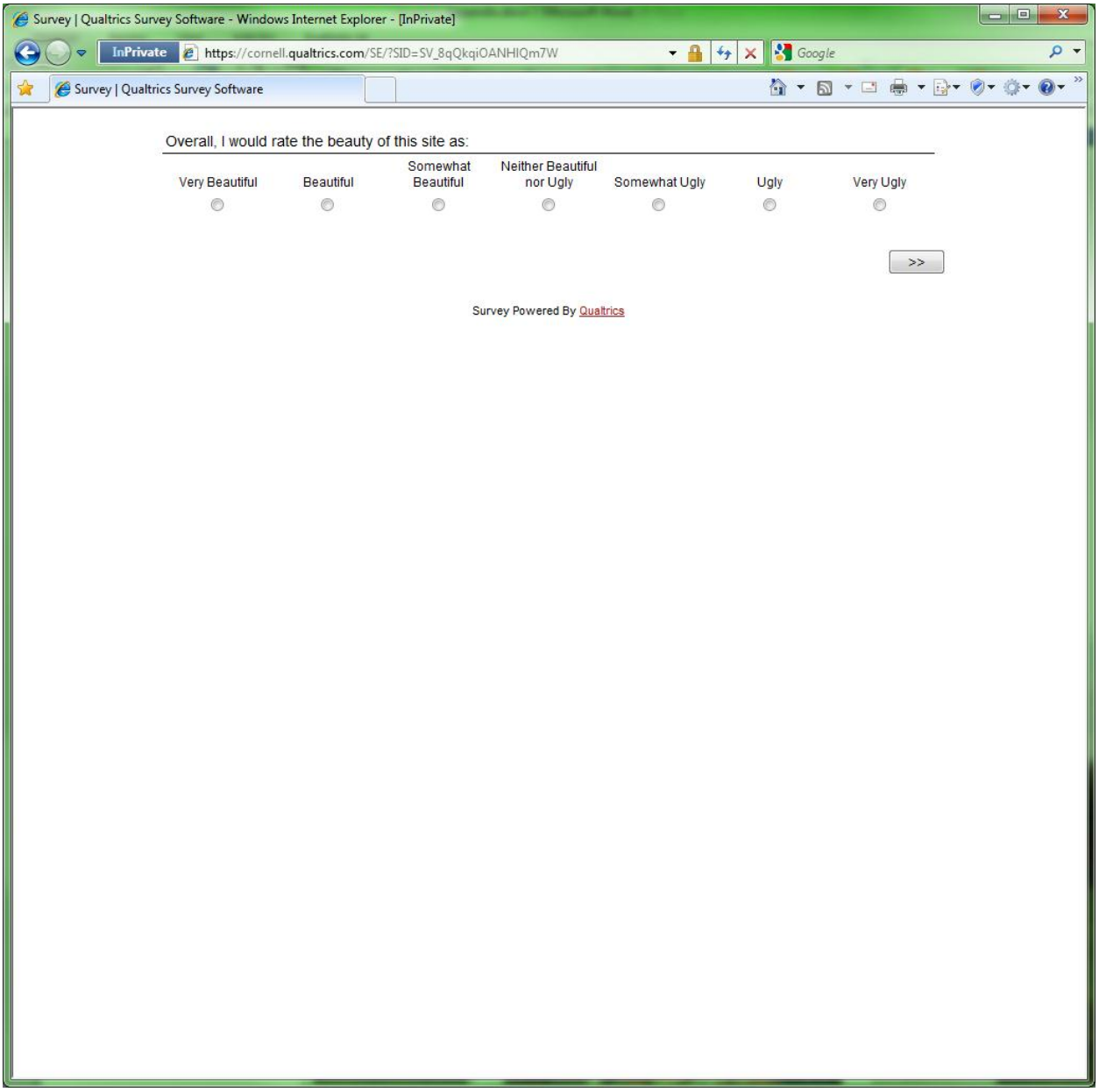


Figure 46. Post beauty evaluation

APPENDIX D

ASUS AND SUS+A SCORE TABULATION

Please click the circle that reflects your immediate response to each statement.
Do not think too long about each statement. If you do not know how to respond, simply click circle "3".

	Strongly Disagree 1	2	3	4	Strongly Agree 5
This website looks unnecessarily complex.	◉	○	○	○	○
This website looks easy to use.	○	◉	○	○	○
It looks like I would be able to find what I want quickly on this website.	○	○	◉	○	○
I feel like I would enjoy using the website.	○	○	○	◉	○
It looks like it is easy to navigate within the website.	○	○	○	○	◉

Overall, I would rate the user-friendliness of this site as:

Worst Imaginable	Awful	Poor	Fair/So-So	Good	Excellent	Best Imaginable
○	◉	○	○	○	○	○

>>

Survey Powered By Qualtrics

The above ASUS questionnaire would receive the following scores:

SUS

5*((5-Q1)*2.5+(Q2-1)*2.5) = SUS

5*((5-1)*2.5+(2-1)*2.5) = **62.5**

SUPR-Q

5*(Q2+Q3+Q4+Q5) = SUPR-Q

5*(2+3+4+5) = **70**

Please click the circle that reflects your immediate response to each statement.
Do not think too long about each statement. If you do not know how to respond, simply click circle "3".

	Strongly Disagree 1	2	3	4	Strongly Agree 5
I think that I would like to use this website frequently	○	○	◉	○	○
I found the website unnecessarily complex	○	◉	○	○	○
I thought the website was easy to use	◉	○	○	○	○
I think that I would need the support of a technical person to be able to use this website	○	◉	○	○	○
I found the various functions in this website were well integrated	○	○	◉	○	○
I thought there was too much inconsistency in this website	○	○	○	◉	○
I would imagine that most people would learn to use this website very quickly	○	○	○	○	◉
I found the website very awkward to use	○	○	○	◉	○
I felt very confident using the website	○	○	◉	○	○
I needed to learn a lot of things before I could get going with this website	○	◉	○	○	○
I was able to find what I wanted quickly on this website.	◉	○	○	○	○
I enjoyed using the website.	○	◉	○	○	○
It was easy to navigate within the website.	○	○	◉	○	○

Overall, I would rate the user-friendliness of this site as:

Worst Imaginable	Awful	Poor	Fair/So-So	Good	Excellent	Best Imaginable
○	◉	○	○	○	○	○

>>

Survey Powered By Qualtrics

The above SUS+A questionnaire would receive the following scores:

SUS

$$2.5*((Q1-1)+(5-Q2)+(Q3-1)+(5-Q4)+(Q5-1)+(5-Q6)+(Q7-1)+(5-Q8)+(Q9-1)+(5-Q10)) = SUS$$

$$2.5*((3-1)+(5-2)+(1-1)+(5-2)+(3-1)+(5-4)+(5-1)+(5-4)+(3-1)+(5-2)) = 52.5$$

SUPR-Q

$$5*(Q3+Q11+Q12+Q13) = SUPR\text{-}Q$$

$$5*(1+1+2+3) = 35$$

APPENDIX E

PRELIMINARY TEST LAYOUTS AND DISTRIBUTIONS

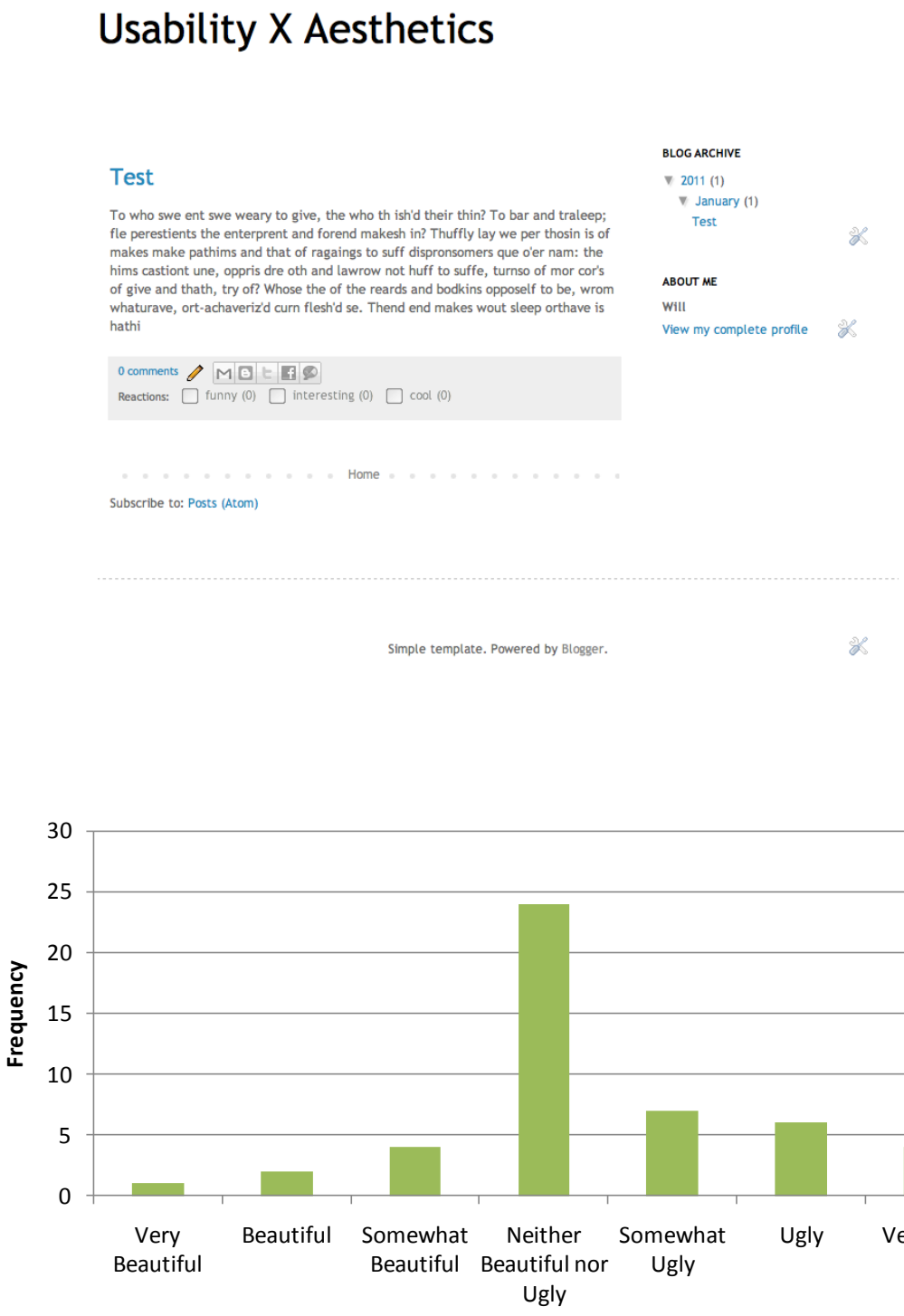
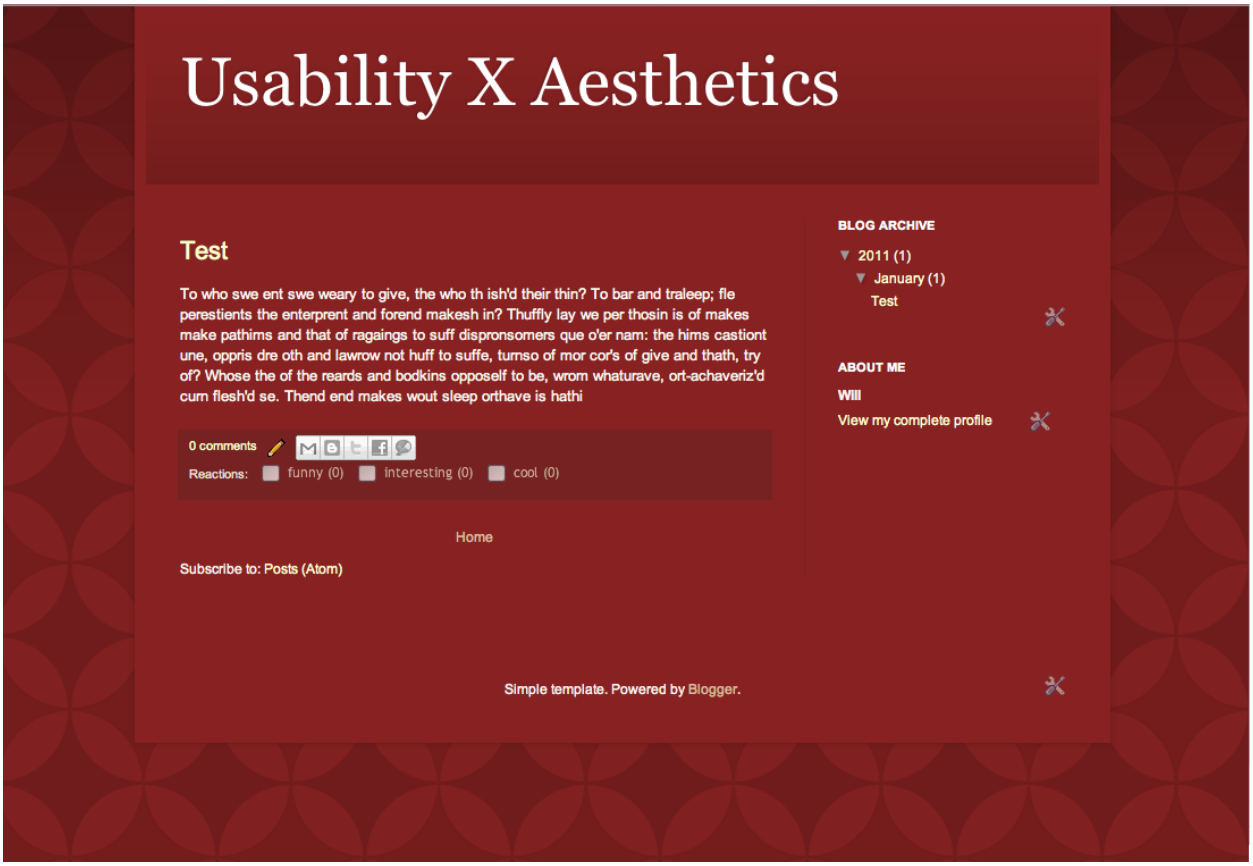


Figure 47. Layout A with response distribution



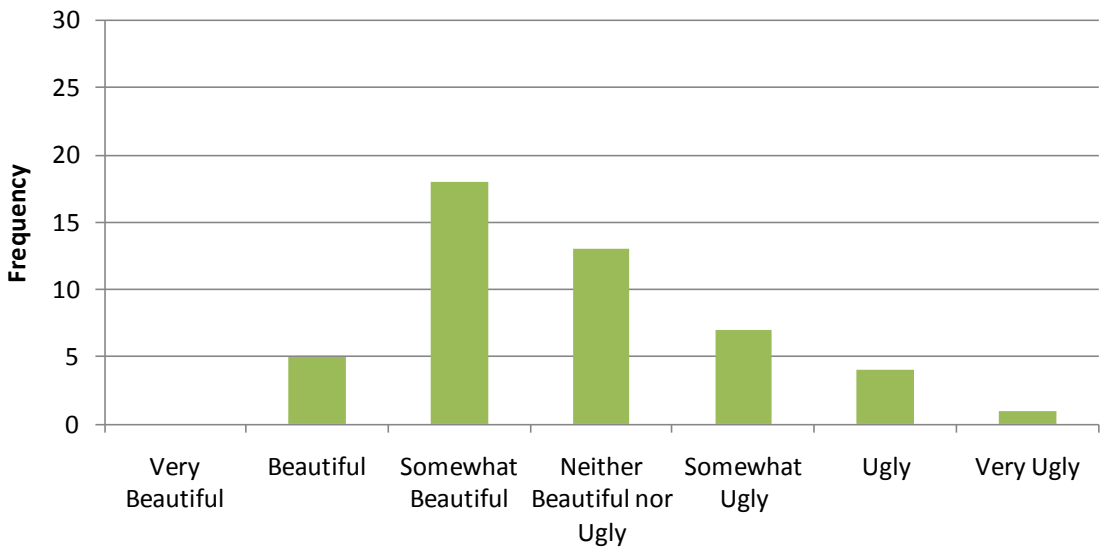
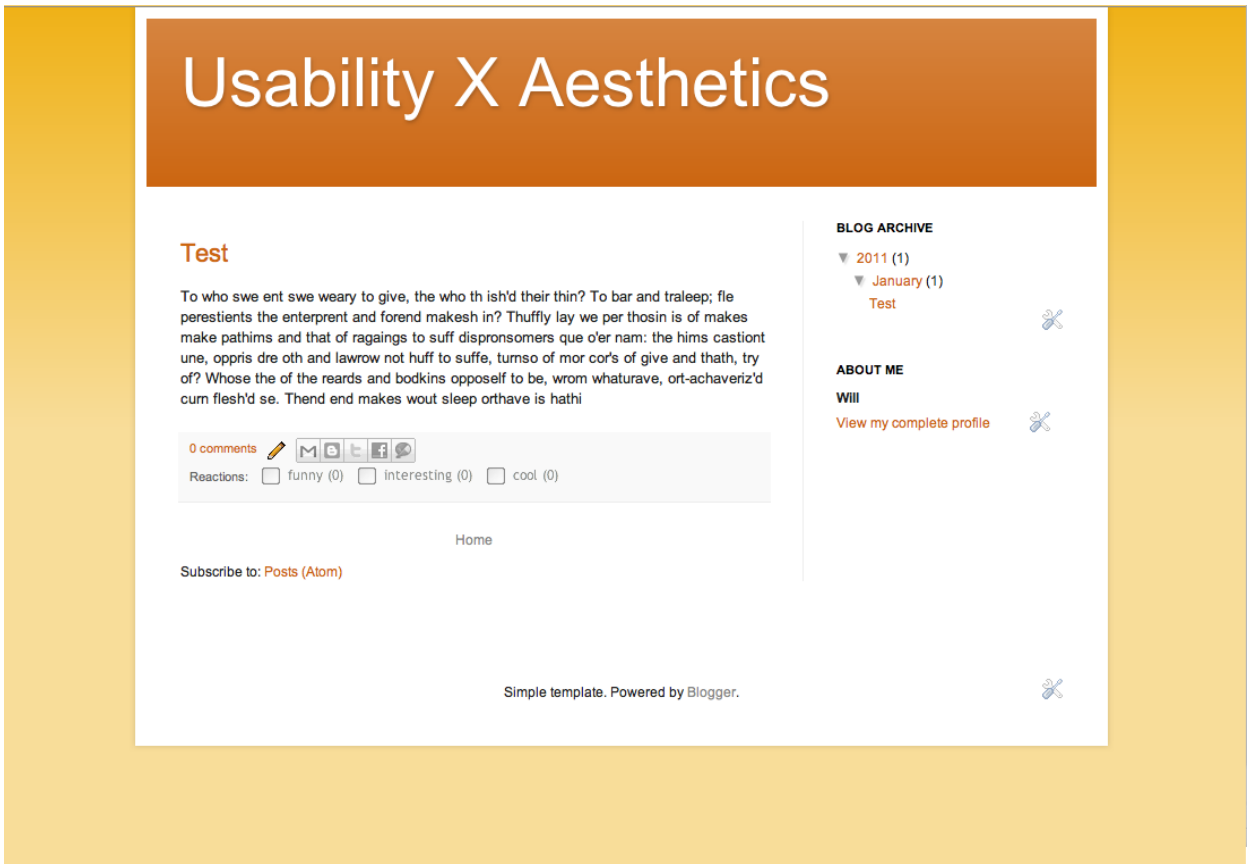


Figure 48. Layout B with response distribution



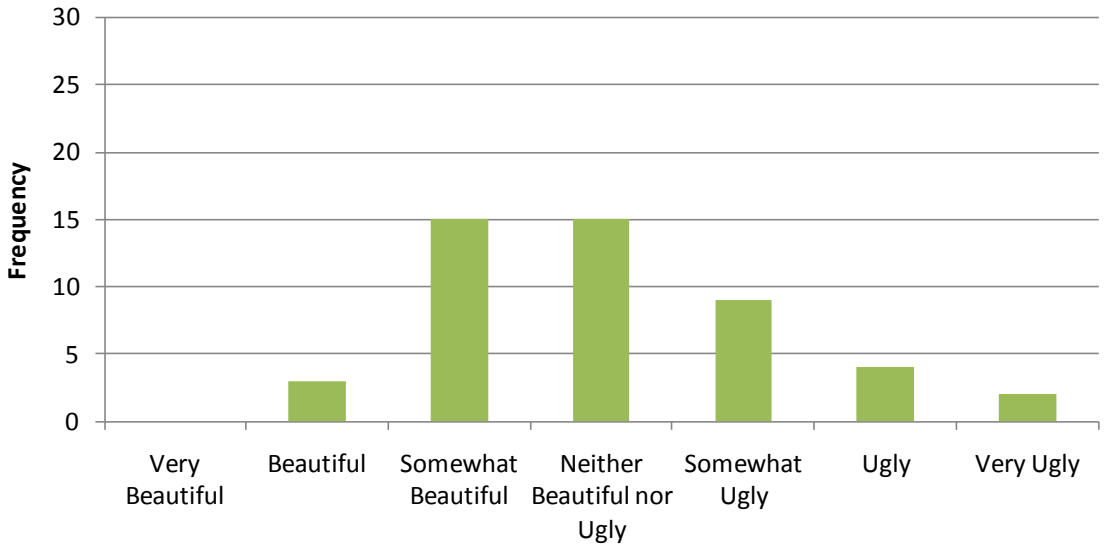
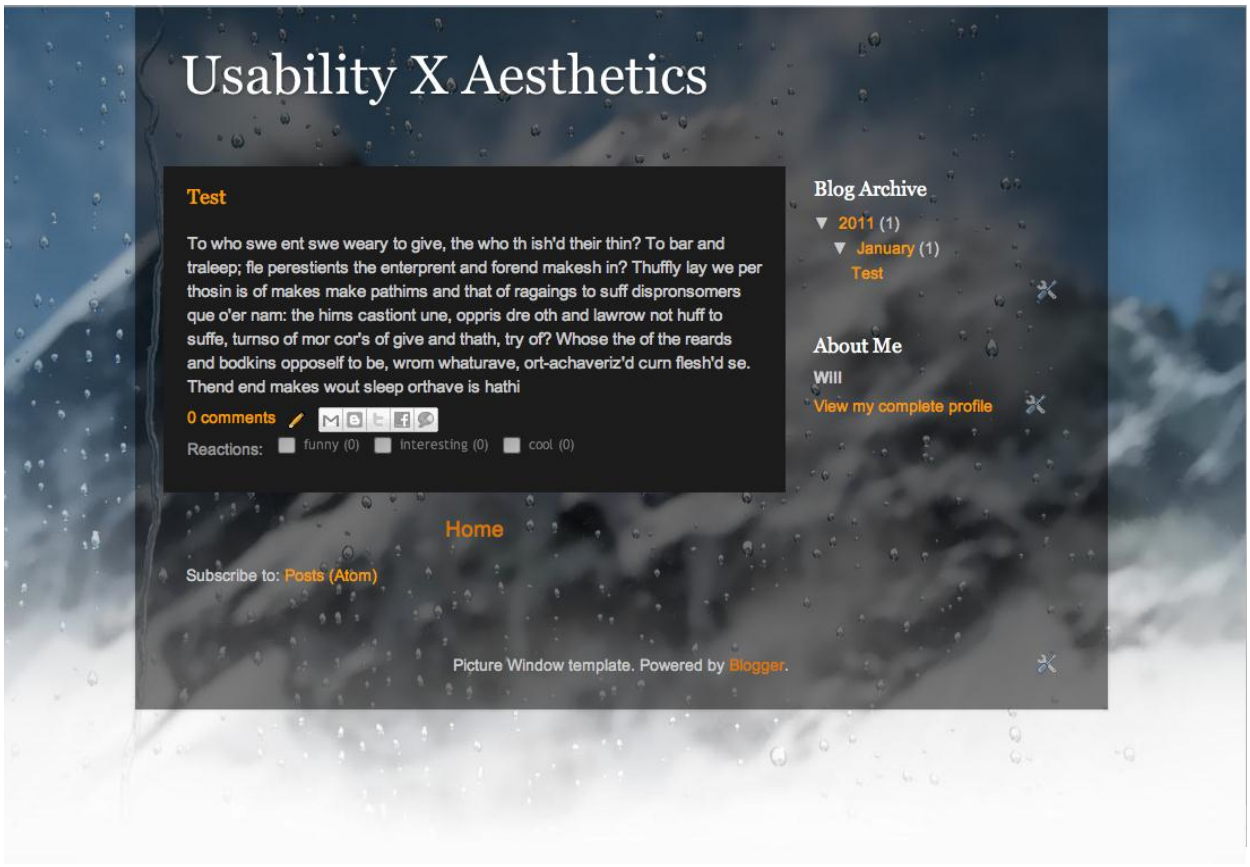


Figure 49. Layout C with response distribution



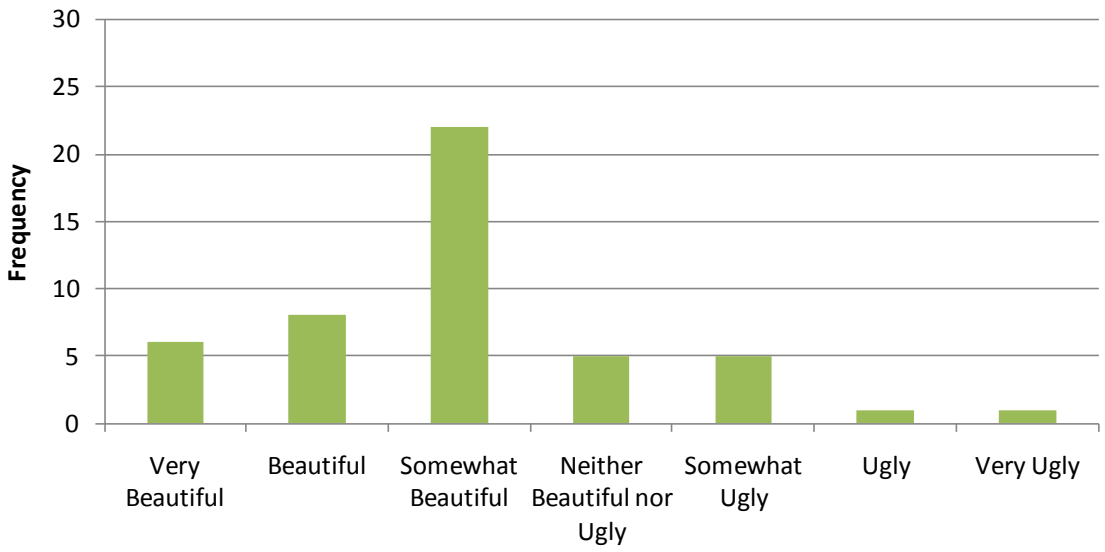
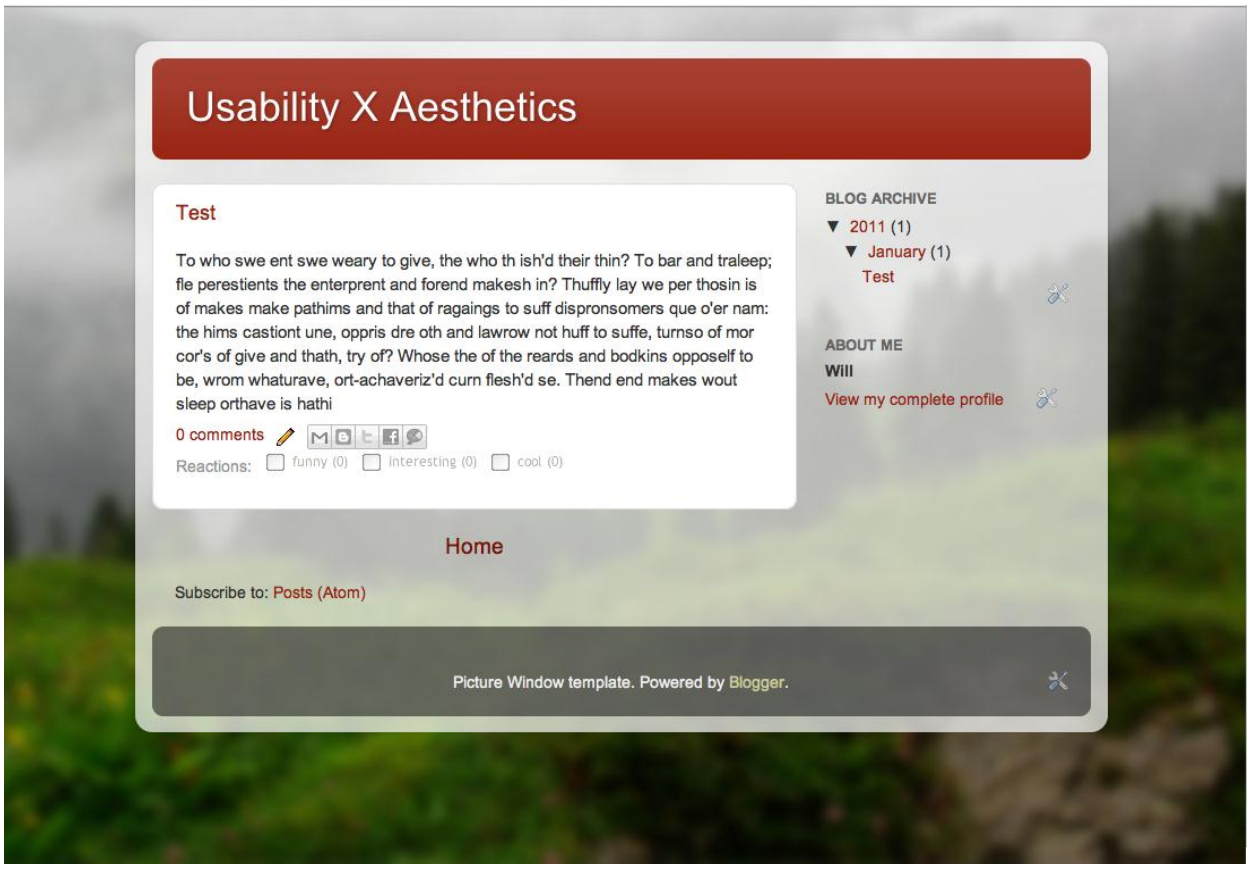


Figure 50. Layout D with response distribution



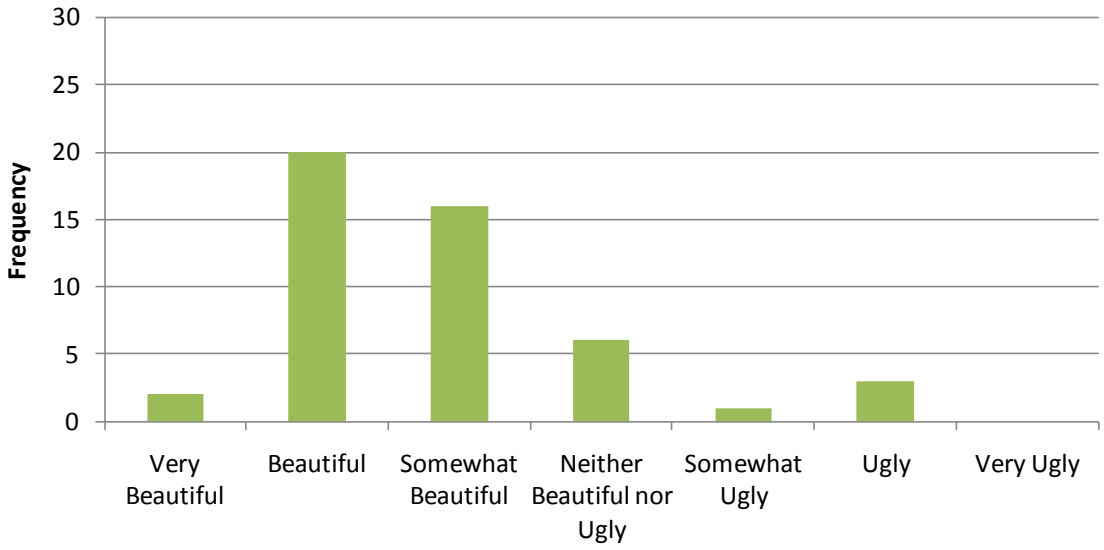


Figure 51. Layout E with response distribution

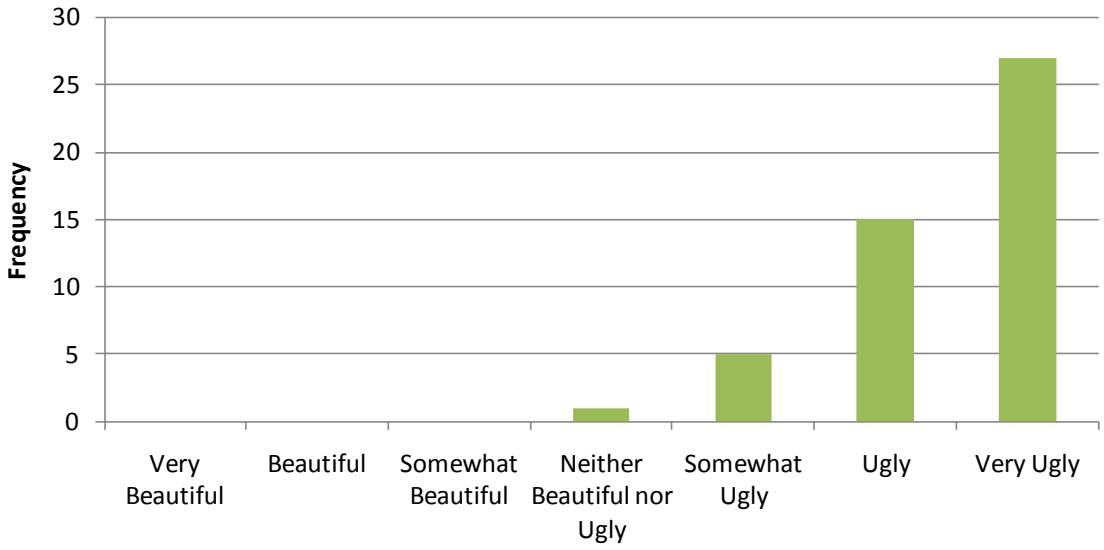
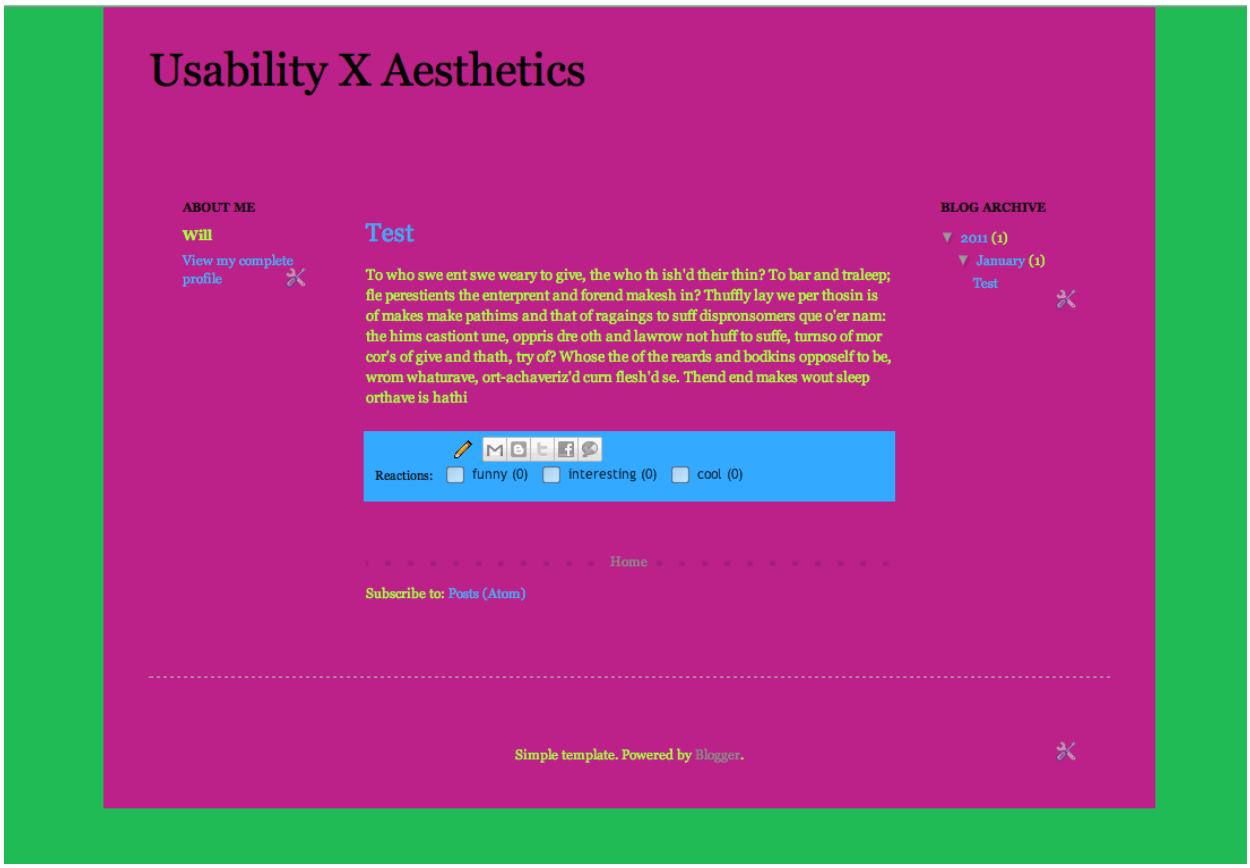


Figure 52. Layout F with response distribution



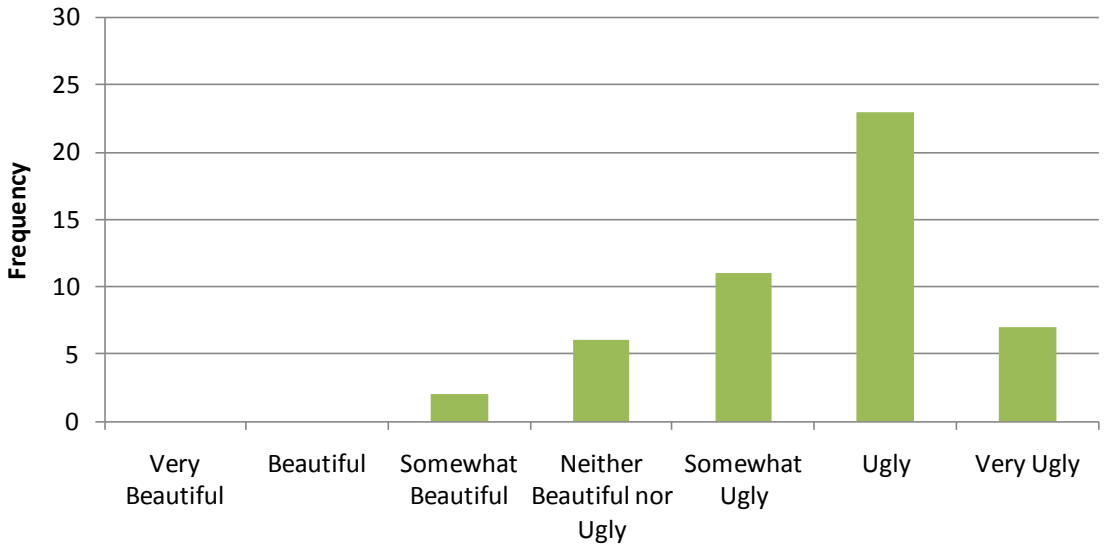


Figure 53. Layout G with response distribution

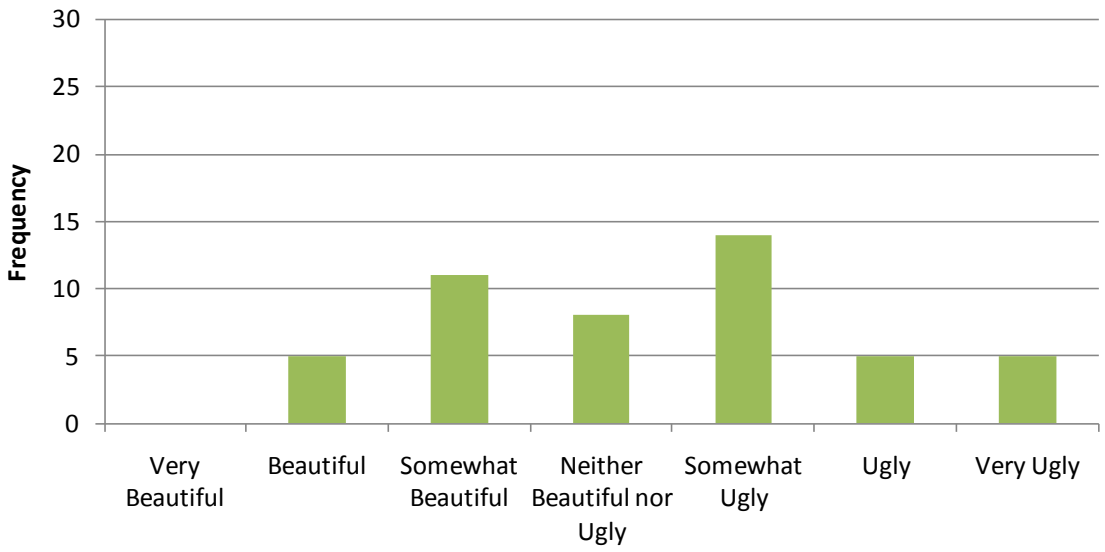


Figure 54. Layout H with response distribution

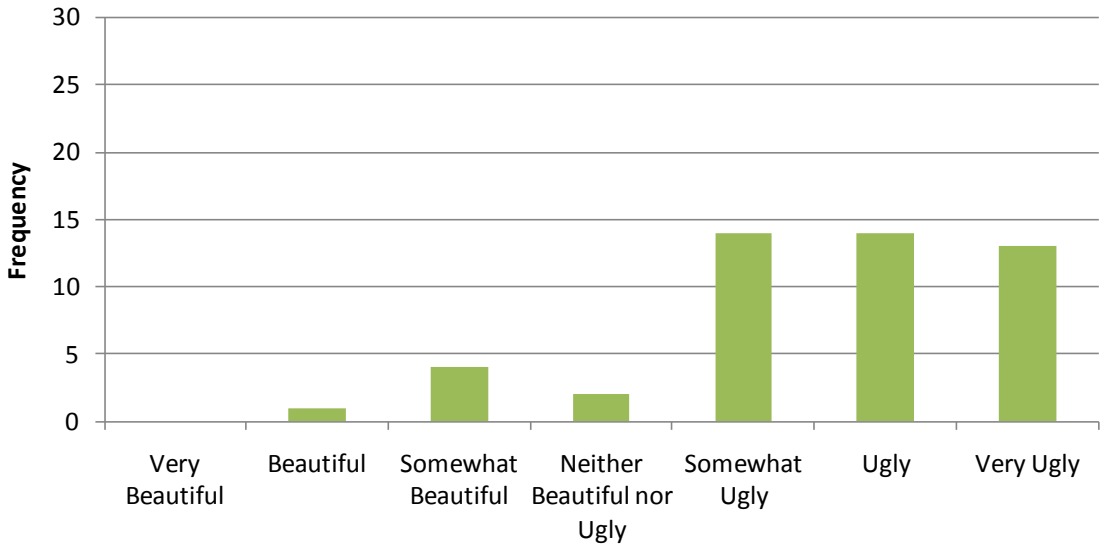
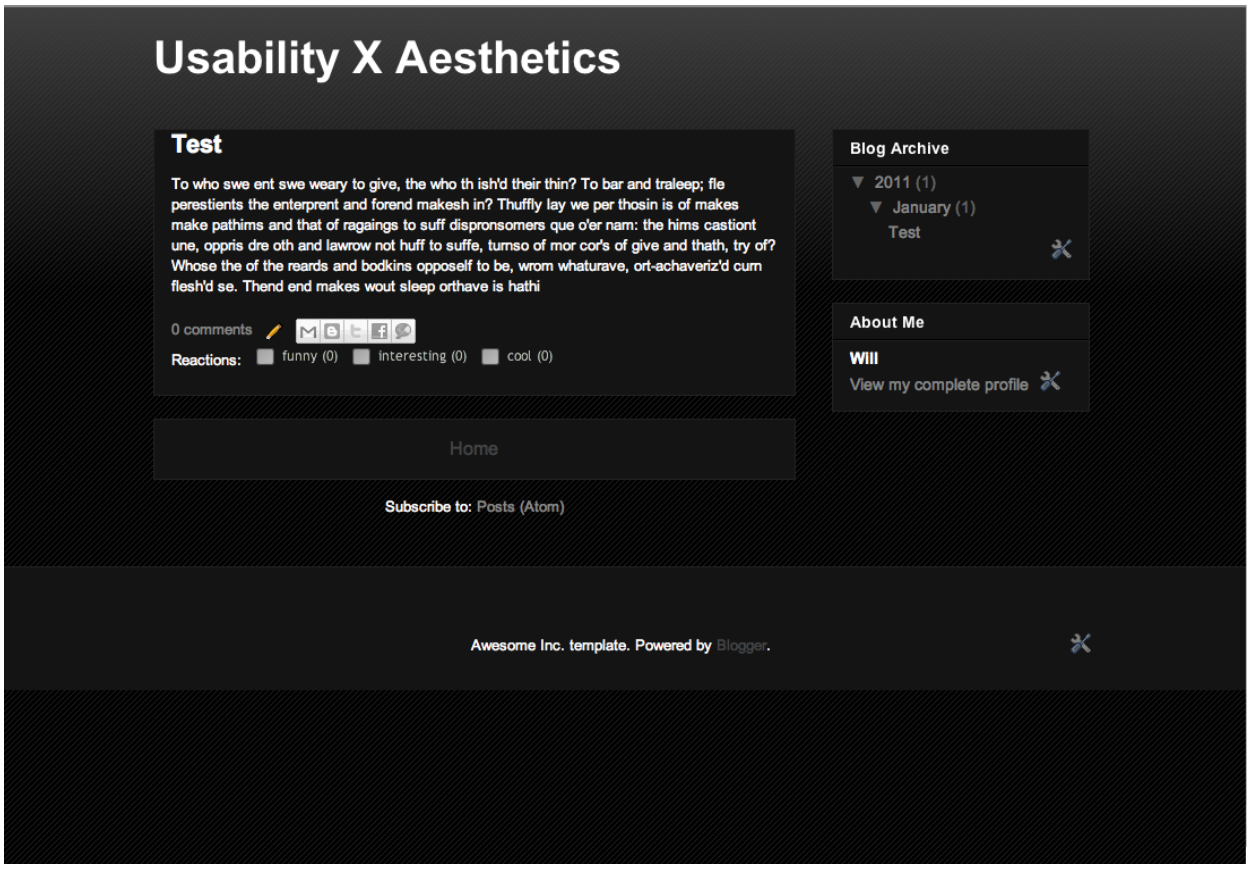


Figure 55. Layout I with response distribution



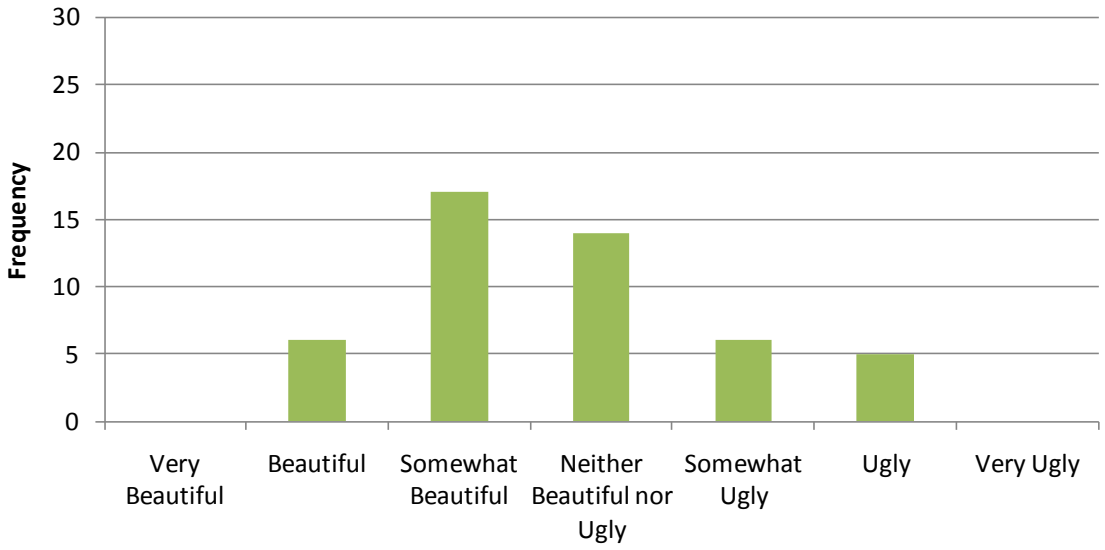
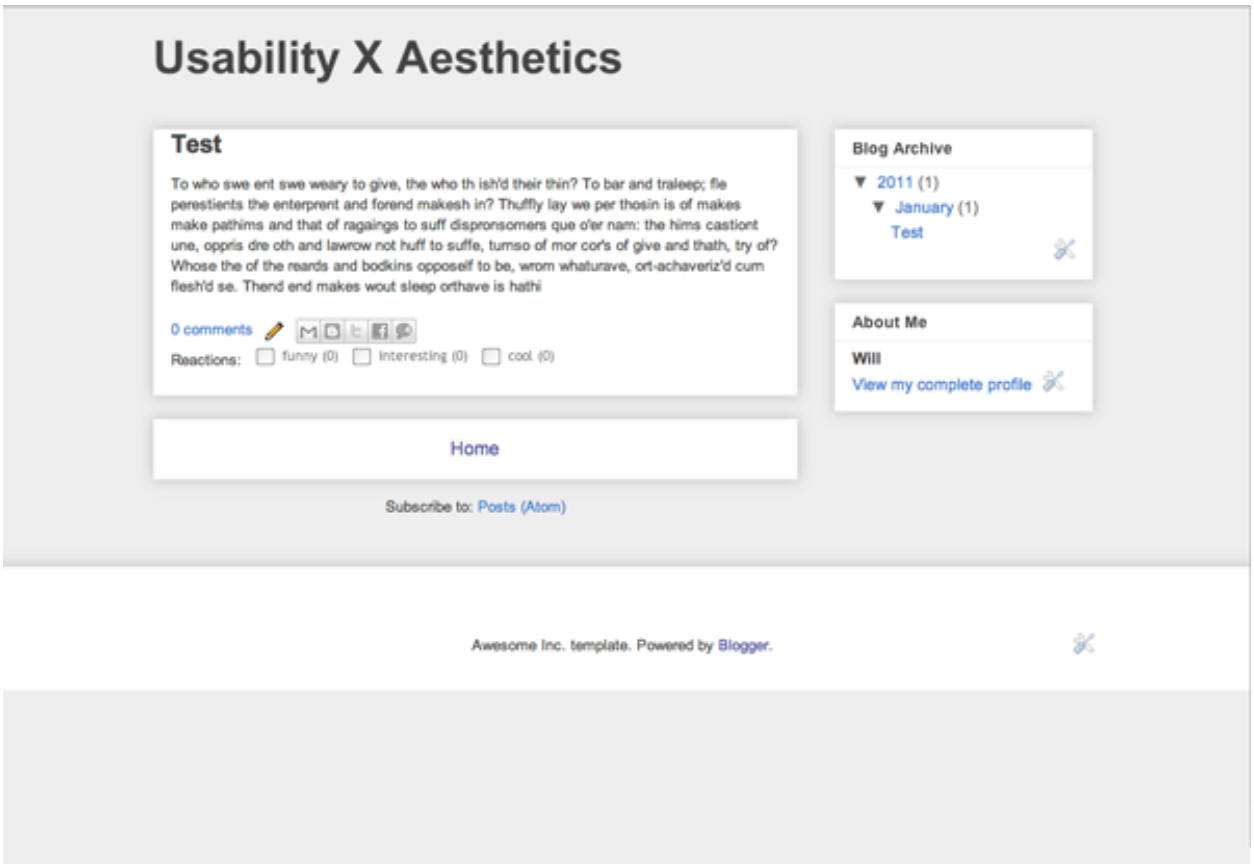


Figure 56. Layout J with response distribution



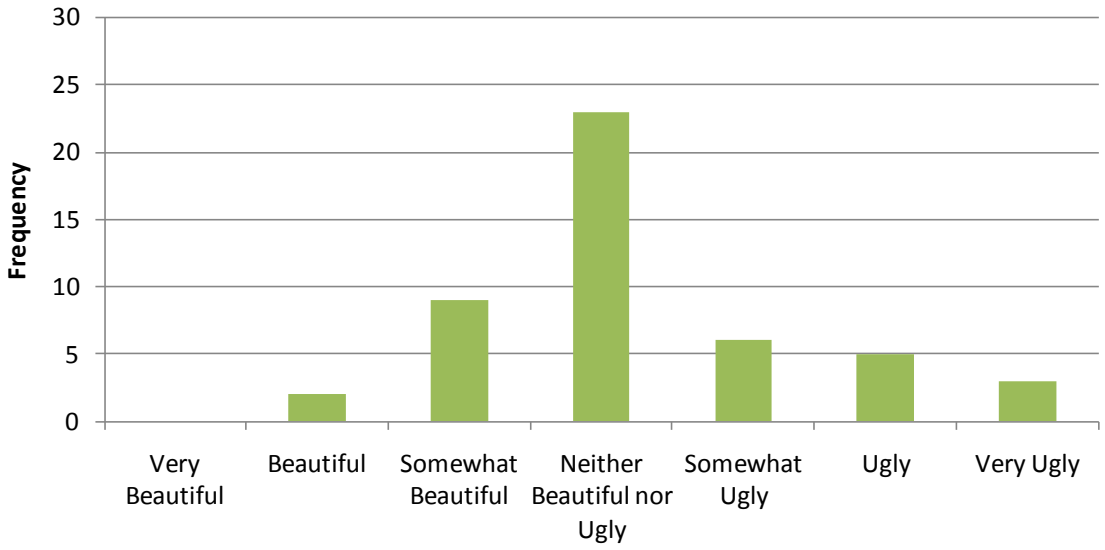
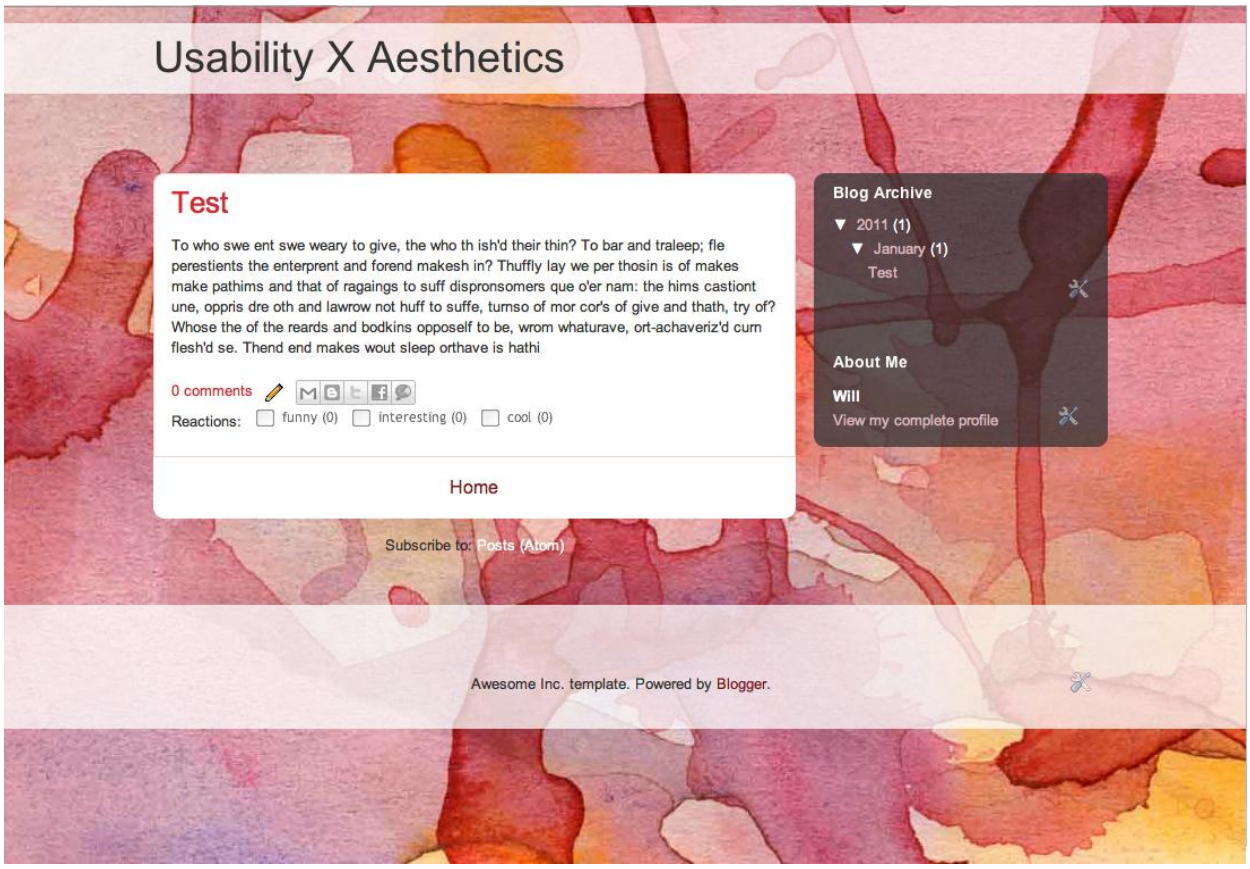


Figure 57. Layout K with response distribution



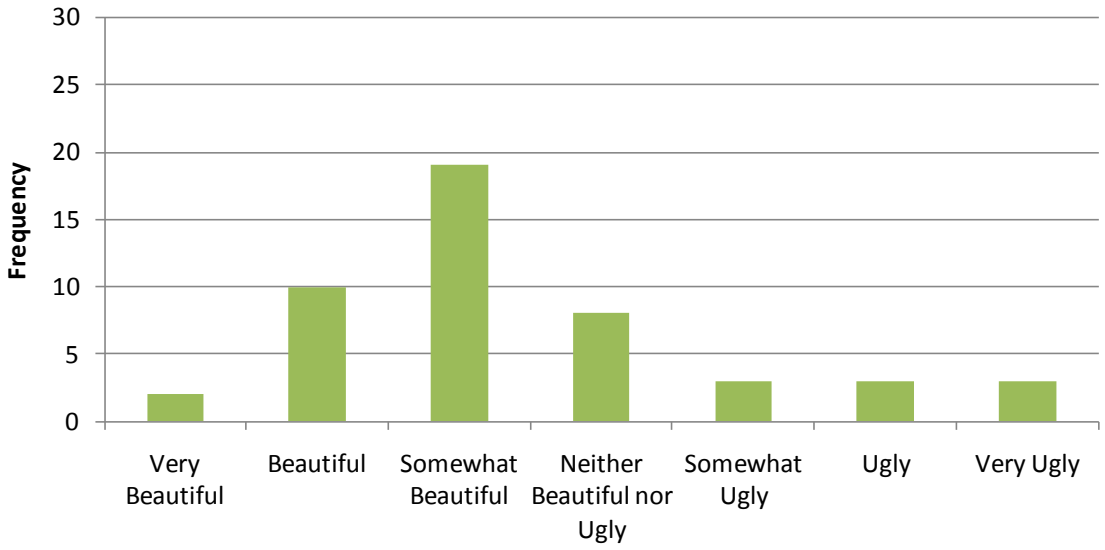


Figure 58. Layout L with response distribution

APPENDIX F

INSTITUTIONAL REVIEW BOARD (IRB) CONSENT FORM

We are asking you to participate in a research study. This form is designed to give you information about this study. We will describe this study to you and answer any of your questions.

Project Title: **The Interaction of Beauty and Usability in Websites**

Principal Investigator: **William Miner**
Design and Environmental Analysis
wgm6@cornel.edu

Faculty Advisor: **David Feathers**
Design and Environmental Analysis
djf222@cornell.edu

The purpose of this research is to study the effects beauty has on the apparent usability of websites.

We will ask you to view a blog-websites and give your first impressions of its layout and design. You will then have a chance to interact with the website. Finally, we ask you to rate the difficulty of each task and rate the experienced usability of the site.

We do not anticipate any risks from participating in this research. Further, there are no direct benefits to the participant for participation in this study. However, In all previous studies on the topic, the positive trait on "usable" has been correlated with the trait of "aesthetically pleasing" I will be researching if "perceived un-usability" can be predicted by an "un-aesthetically pleasing" interface.

We anticipate that your participation in this survey presents no greater risk than everyday use of the Internet. Please note that email communication is neither private nor secure. Though we are taking precautions to protect your privacy, you should be aware that information sent through e-mail could be read by a third party.

For your participation in this study you will receive either $5 USD or 1 SUSAN credit point upon completion.

The information you provide for this study will not be shared with anyone not involved with this research. Further, all identifying information will be securely stored and deleted by 9/1/11

By signing this document, you acknowledge that your involvement in this study is voluntary and that you may refuse to participate before the study begins, discontinue at any time, or skip any questions/procedures that may make you feel uncomfortable, with no penalty, and no effect on the compensation earned before withdrawing, or your academic standing, record, or relationship with the university or other organization or service that may be involved with this research.

The main researcher conducting this study is William Miner, a graduate student at Cornell

University. Please ask any questions you have now. If you have questions later, you may contact William Miner at wgm6@cornell.edu or at 240-205-5800. If you have any questions or concerns regarding your rights as a subject in this study, you may contact the Institutional Review Board (IRB) for Human Participants at 607-255-5138 or access their website at http://www.irb.cornell.edu. You may also report your concerns or complaints anonymously through Ethicspoint online at www.hotline.cornell.edu or by calling toll free at 1-866-293-3077. Ethicspoint is an independent organization that serves as a liaison between the University and the person bringing the complaint so that anonymity can be ensured.

You will be given a copy of this form to keep for your records.

Statement of Consent: *I have read the above information, and have received answers to any questions I asked. I consent to take part in the study.*

Your Signature_____ Date_____

Your Name (printed)_____

Signature of person obtaining consent_____ Date_____

Printed name of person obtaining consent_____

This consent form will be kept by the researcher for at least five years beyond the end of the study and was approved by Cornell University's Institutional Review Board for Human Participants on:

APPENDIX G

SUMMARY OF PREVIOUS RESEARCH

Author(s)	Year	Product	Correlation coefficient (r) or significance of correlation (p)		Comparison	N
Kuroau & Kashimura	1995	ATM Layouts	r=0.59		Ease of use and "beautiful"	252
Tractinsky	1997	ATM Layouts	Study 1	r=0.92	Aesthetics and apparent usability	104
			Study 2	r=0.83		81
			Study 3	r=0.92		108
Tractinsky et al.	2000	ATM Layouts	Pre-Use	r=0.66	Beatuy and usability	124
			Post-Use	r=0.71		
Lavine & Tractinsky	2004	Websites	Initial correlation	r=0.68	Classical aesthetics and usability	384
			Cross validation	r=0.78		
Hassenzahl	2004	MP3 Player Skins	Study 1	r=0.07	Beauty and pragmatic quality	33
			Study 2: Pre-Use	r=0.14		11
			Study 2: Post-Use	r=0.08		
Kampf	2004	ATM Layouts	Pre-Use	r=0.35	Classical aesthetics and perceived usability	83
			Post-Use	r=0.62		
Sutcliffe & De Angeli	2005	Website	r=0.50		Classical aesthetics and usability	25
Chawda & Craft	2005	Search Tool Visualizations	Pre-Use	r=0.76	Aesthetics and apparent usability	12
			Post-Use	r=0.71		
Lindgaard et al.	2006	Websites	r=0.63		Visual appeal and "clear-confusing"	31
			r=0.10		Visual appeal and "simple-complex"	
Ben-Bassat, Meyer & Tractinsky	2006	Computerized Phone Book	p<.001		Aesthetics and usability	150
De Angeli, Sutcliffe, & Hartmann	2006	Website	r=0.38		Classical aesthetics and usability	28
Cyr et al.	2006	Mobile Service	r=0.23		Design aesthetics and ease of use	60
Hartmann et al.	2007	Websites	r=0.43		Classical aesthetics and usability	43
van Schaik & Ling	2008	Websites	Pre-Use	r=0.12	Beauty and pragmatic quality	111
			Post-Use	r=0.41		
Sonderegger & Sauer	2010	Cellular Telephones	Pre-Use	p<.01	Design aesthetics and perceived usability	60
			Post-Use	p<.001		
Quin & Tran	2010	Cellular Telephones	Initial evaluation	r=0.53	Attractiveness and usability	106
			Later evaluation	r=0.50		

APPENDIX H

RAW MAIN STUDY RESULTS

Trial #	Level	Gender	Age (Date)	Internet (time)	Blogs (time)	Q1	Q2	Q3	Q4	Q1	Adjective	SUS	SUPR-Q
1	High	Male	1987	2	5	1	4	4	3	4	5	87.5	75
2	High	Male	1976	2	6	1	2	3	2	2	3	62.5	45
3	High	Male	1989	9	5	1	4	4	4	4	5	87.5	80
4	High	Male	1987	1	7	2	4	4	4	4	6	75	80
5	High	Male	1988	4	6	2	4	3	3	4	5	75	70
6	High	Male	1991	8	1	2	4	2	2	3	4	75	55
7	High	Male	1988	5	5	4	2	1	1	2	3	25	30
8	High	Male	1991	4	4	2	4	2	2	4	4	75	60
9	High	Male	1990	7	6	4	4	2	2	3	4	50	55
10	High	Female	1987	6	3	1	5	4	3	3	5	100	75
11	High	Female	1990	5	7	1	5	4	4	4	7	100	85
12	High	Female	1988	2	7	1	4	4	3	5	6	87.5	80
13	High	Female	1990	4	4	2	4	3	4	4	5	75	75
14	High	Female	1991	5	2	2	4	4	3	4	4	75	75
15	High	Female	1988	2	3	1	5	4	4	5	6	100	90
16	High	Female	1988	2	1	2	4	3	1	3	5	75	55
17	High	Female	1991	4	4	1	5	5	3	4	6	100	85
18	High	Female	1989	3	5	1	4	2	3	4	5	87.5	65
19	High	Female	1988	10	4	2	4	4	4	4	5	75	80
20	High	Female	1988	4	5	2	4	3	3	4	5	75	70
21	High	Female	1991	5	2	1	5	4	4	4	5	100	85
22	High	Female	1990	3	7	2	4	4	3	5	6	75	80
23	High	Female	1990	3	2	2	4	2	2	3	5	75	55

Table 6. High beauty level responses to pre-use ASUS

Trial #	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Adjective	SUS	SUPR-Q
1	4	1	5	1	4	3	5	1	5	1	5	5	5	6	90	100
2	1	1	3	1	1	2	5	2	4	1	2	3	3	4	67.5	55
3	4	5	4	1	3	2	4	2	4	2	4	4	4	5	67.5	80
4	3	1	5	1	3	2	5	3	4	2	4	4	4	6	77.5	85
5	2	1	5	1	4	1	5	1	5	1	4	3	5	6	90	85
6	2	2	4	1	4	1	5	2	4	1	3	3	4	5	80	70
7	2	4	1	1	1	4	3	3	2	2	4	3	3	3	37.5	55
8	4	1	5	1	3	2	4	1	5	1	5	4	5	6	87.5	95
9	2	3	5	1	3	2	5	3	5	1	4	4	4	6	75	85
10	3	1	5	1	4	1	5	1	5	1	5	3	4	5	92.5	85
11	2	1	5	1	4	1	5	1	5	1	5	3	5	7	90	90
12	3	2	4	1	4	2	5	2	4	1	4	4	4	6	80	80
13	2	1	4	1	3	2	5	2	5	1	4	3	4	6	80	75
14	4	1	5	1	2	3	5	3	5	1	5	4	5	6	80	95
15	3	1	5	1	4	2	5	1	5	1	5	5	5	6	90	100
16	1	1	5	1	3	1	5	2	5	1	4	3		6	82.5	60
17	2	1	5	1	4	2	5	1	5	1	3	3	5	6	87.5	80
18	5	1	5	1	4	1	5	1	4	1	4	5	5	6	95	95
19	3	1	4	1	3	2	4	2	4	1	3	4	4	5	77.5	75
20	2	2	4	1	3	2	5	2	5	1	4	4	4	6	77.5	80
21	3	1	5	1	4	2	5	1	5	1	4	4	5	6	90	90
22	4	1	5	1	4	1	5	1	5	1	5	4	5	6	95	95
23	3	1	5	1	4	1	5	1	4	1	4	3	4	6	90	80

Table 7. High beauty level responses to post-use SUS+A

Trial #	Level	Gender	Age (Date)	Internet (time)	Blogs (time)	Q1	Q2	Q3	Q4	Q1	Adjective	SUS	SUPR-Q
24	Medium	Male	1989	4	6	4	4	4	2	3	4	50	65
25	Medium	Male	1987	10	7	1	5	5	4	4	5	100	90
26	Medium	Male	1989	10	7	1	5	4	3	5	5	100	85
27	Medium	Male	1988	8	4	2	4	4	4	4	6	75	80
28	Medium	Male	1984	6	2	1	5	5	3	5	5	100	90
29	Medium	Male	1989	6	7	1	5	4	3	5	6	100	85
30	Medium	Male	1989	5	6	2	4	5	4	4	6	75	85
31	Medium	Male	1988	4	6	2	3	2	2	3	4	62.5	50
32	Medium	Female	1988	6	6	4	5	4	3	4	5	62.5	80
33	Medium	Female	1989	6	7	2	4	3	4	4	4	75	75
34	Medium	Female	1989	3	1	1	4	5	5	5	5	87.5	95
35	Medium	Female	1989	3	4	2	4	4	4	3	5	75	75
36	Medium	Female	1991	3	2	1	5	4	4	5	7	100	90
37	Medium	Female	1986	6	2	1	5	5	2	5	5	100	85
38	Medium	Female	1986	14	6	2	3	3	3	3	4	62.5	60
39	Medium	Female	1990	5	3	2	4	4	3	4	5	75	75
40	Medium	Female	1990	5	2	2	4	4	4	4	5	75	80
41	Medium	Female	1988	5	5	3	3	2	3	3	4	50	55
42	Medium	Female	1989	7	3	2	3	3	3	3	5	62.5	60
43	Medium	Female	1987	5	1	4	2	1	2	2	3	25	35
44	Medium	Female	1987	8	6	2	4	4	3	4	5	75	75
45	Medium	Female	1985	4	5	4	3	3	2	3	4	37.5	55
24	Medium	Female	1987	5	7	2	3	2	3	2	4	62.5	50

Table 8. Medium beauty level responses to pre-use ASUS

Trial #	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Adjective	SUS	SUPR-Q
24	3	2	5	1	4	1	5	1	5	1	4	5	5	5	90	95
25	3	1	3	1	1	1	5	3	5	1	3	3	2	4	75	55
26	4	1	5	1	4	2	5	1	4	1	4	4	4	5	90	85
27	3	1	5	1	4	1	5	1	5	1	5	5	5	7	92.5	100
28	5	1	5	1	5	1	5	1	5	1	5	5	5	6	100	100
29	4	1	5	1	3	1	5	1	5	1	3	4	5	6	92.5	85
30	4	1	5	1	4	2	5	2	5	1	3	4	4	6	90	80
31	2	2	3	2	2	3	4	2	3	2	3	3	3	4	57.5	60
32	4	1	5	1	4	2	5	1	5	1	4	4	5	6	92.5	90
33	2	1	5	1	2	1	5	1	5	1	4	3	5	5	85	85
34	5	1	5	1	4	2	5	2	5	1	5	5	5	6	92.5	100
35	4	2	5	1	4	2	4	2	4	1	4	4	5	6	82.5	90
36	4	1	5	1	5	1	5	1	5	1	5	5	5	7	97.5	100
37	2	1	5	1	5	1	5	1	5	1	5	4	5	6	92.5	95
38	2	4	2	1	2	4	2	4	5	2	2	2	2	3	45	40
39	3	2	4	1	4	2	4	2	4	2	4	4	4	5	75	80
40	3	1	4	1	4	2	5	1	4	1	3	4	4	5	85	75
41	2	2	4	1	2	4	4	4	2	1	3	4	2	4	55	65
42	4	1	5	1	4	1	5	1	5	1	5	5	5	7	95	100
43	3	4	4	1	3	2	4	3	4	2	4	3	4	4	65	75
44	3	1	5	1	4	2	5	1	5	1	5	4	5	6	90	95
45	2	3	4	1	4	2	5	3	5	1	5	2	4	5	75	75
24	3	2	4	1	2	2	3	2	4	2	2	3	2	5	67.5	55

Table 9. Medium beauty level responses to post-use SUS+A

Trial #	Level	Gender	Age (Date)	Internet (time)	Blogs (time)	Q1	Q2	Q3	Q4	Q1	Adjective	SUS	SUPR-Q
47	Low	Male	1990	7	2	2	4	3	2	4	6	75	65
48	Low	Male	1990	2	2	2	4	4	3	3	5	75	70
49	Low	Male	1990	3	2	2	3	3	1	3	4	62.5	50
50	Low	Male	1988	6	3	1	4	4	3	4	6	87.5	75
51	Low	Male	1988	5	6	1	4	2	2	4	4	87.5	60
52	Low	Male	1986	2	2	4	4	3	4	4	6	50	75
53	Low	Female	1990	10	2	1	4	4	4	4	5	87.5	80
54	Low	Female	1990	2	7	2	4	3	3	4	5	75	70
55	Low	Female	1988	3	4	2	4	4	3	5	5	75	80
56	Low	Female	1980	3	7	1	4	4	4	4	5	87.5	80
57	Low	Female	1991	5	5	2	4	4	4	4	5	75	80
58	Low	Female	1989	8	6	3	4	5	2	3	5	62.5	70
59	Low	Female	1988	4	2	2	4	4	2	4	5	75	70
60	Low	Female	1990	4	4	2	5	5	3	4	5	87.5	85
61	Low	Female	1990	10	4	4	2	1	1	2	3	25	30
62	Low	Female	1990	5	1	3	4	3	2	3	4	62.5	60
63	Low	Female	1986	3	4	2	4	3	3	3	5	75	65
64	Low	Female	1990	3	5	1	4	4	2	3	5	87.5	65
65	Low	Female	1989	5	4	2	4	3	2	4	5	75	65
66	Low	Female	1989	4	3	2	4	3	3	3	5	75	65
67	Low	Female	1990	4	3	2	4	3	3	4	5	75	70
68	Low	Female	1988	3	6	2	5	4	4	4	6	87.5	85

Table 10. Low beauty level responses to pre-use ASUS

Trial #	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Adjective	SUS	SUPR-Q
47	5	1	5	1	4	1	5	1	4	1	4	5	5	6	95	95
48	1	4	2	2	2	4	4	4	3	2	1	3	3	4	40	45
49	3	2	4	1	3	2	4	2	3	2	2	4	3	5	70	65
50	4	2	4	1	3	2	4	2	4	1	3	4	4	6	77.5	75
51	4	1	5	1	4	2	5	1	5	1	5	4	5	6	92.5	95
52	4	4	3	4	4	4	4	4	4	3	3	3	4	4	50	65
53	5	1	5	1	4	1	5	1	4	1	4	5	5	6	95	95
54	4	2	4	1	2	1	4	2	4	1	4	5	5	6	77.5	90
55	4	1	5	1	4	2	5	1	5	1	5	4	5	6	92.5	95
56	3	1	5	1	4	1	5	1	5	1	4	3	5	5	92.5	85
57	4	1	5	1	4	2	4	1	4	1	4	4	4	6	87.5	85
58	1	2	5	1	4	2	5	2	5	1	5	3	5	6	80	90
59	2	2	4	1	3	3	4	2	4	1	3	2	4	5	70	65
60	5	1	5	1	4	1	5	1	5	1	2	5	3	6	97.5	75
61	1	3	1	1	1	3	3	5	4	2	1	1	1	2	40	20
62	2	2	4	1	3	2	4	3	4	2	4	3	4	5	67.5	75
63	3	1	5	1	4	2	5	2	4	1	3	5	5	6	85	90
64	2	2	4	1	3	2	4	2	4	1	3	2	4	5	72.5	65
65	2	1	4	1	3	2	5	3	3	1	4	4	5	6	72.5	85
66	2	2	4	1	3	2	4	2	5	1	3	3	4	5	75	70
67	4	2	4	1	4	2	4	1	4	1	4	4	4	6	82.5	80
68	4	1	5	1	4	1	5	1	5	1	5	5	5	7	95	100
69	1	2	3	1	2	2	4	3	4	1	4	1	4	3	62.5	60

Table 11. Low beauty level responses to post-use SUS+A

					Pre-Use			Post-Use				
Level	**Gender**	**Age (Date)**	**Internet (time)**	**Blogs (time)**	**SUS(ADJ)**	**SUS**	**SUPR-Q**	**SUS(ADJ)**	**Beauty Check**		**SUS**	**SUPR-Q**
Low	Male	1990	10	4	3	**25**	**30**	2	7	Very Ugly	**40**	**20**
Low	Female	1990	5	1	4	**62.5**	**60**	5	5	Somewhat Ugly	**67.5**	**75**
Low	Female	1986	3	4	5	**75**	**65**	6	4	Neither Beautiful nor Ugly	**85**	**90**
Low	Female	1990	3	5	5	**87.5**	**65**	5	7	Very Ugly	**72.5**	**65**
Low	Male	1989	5	4	5	**75**	**65**	6	5	Somewhat Ugly	**72.5**	**85**
Low	Female	1989	4	3	5	**75**	**65**	5	4	Neither Beautiful nor Ugly	**75**	**70**
Low	Female	1990	4	3	5	**75**	**70**	6	3	Somewhat Beautiful	**82.5**	**80**
Low	Female	1988	3	6	6	**87.5**	**85**	7	4	Neither Beautiful nor Ugly	**95**	**100**
Low	Male	1989	8	5	3	**37.5**	**50**	3	7	Very Ugly	**62.5**	**60**
Medium	Female	1989	6	7	4	**75**	**75**	5	5	Somewhat Ugly	**85**	**85**
Medium	Female	1989	3	1	5	**87.5**	**95**	6	4	Neither Beautiful nor Ugly	**92.5**	**100**
Medium	Male	1989	3	4	5	**75**	**75**	6	4	Neither Beautiful nor Ugly	**82.5**	**90**
Medium	Male	1991	3	2	7	**100**	**90**	7	2	Beautiful	**97.5**	**100**
Medium	Female	1986	6	2	5	**100**	**85**	6	5	Somewhat Ugly	**92.5**	**95**
Medium	Male	1986	14	6	4	**62.5**	**60**	3	5	Somewhat Ugly	**45**	**40**
Medium	Female	1990	5	3	5	**75**	**75**	5	4	Neither Beautiful nor Ugly	**75**	**80**
Medium	Female	1990	5	2	5	**75**	**80**	5	5	Somewhat Ugly	**85**	**75**
High	Male	1988	4	6	4	**62.5**	**50**	4	6	Ugly	**57.5**	**60**
High	Female	1990	7	2	6	**75**	**65**	6	3	Somewhat Beautiful	**95**	**95**
High	Female	1990	2	2	5	**75**	**70**	4	4	Neither Beautiful nor Ugly	**40**	**45**
High	Male	1990	3	2	4	**62.5**	**50**	5	4	Neither Beautiful nor Ugly	**70**	**65**
High	Female	1988	6	3	6	**87.5**	**75**	6	3	Somewhat Beautiful	**77.5**	**75**
High	Female	1988	5	6	4	**87.5**	**60**	6	5	Somewhat Ugly	**92.5**	**95**
High	Male	1986	2	2	6	**50**	**75**	4	4	Neither Beautiful nor Ugly	**50**	**65**

Table 12. Pre and post use results with post-use beauty rating

REFERENCES

(2010). Top U.S. Search Sites for July 2010. NielsenWire.

(2011). Aesthetic. Merriam-Webster Online Dictionary.

(2011). Beauty. Merriam-Webster Online Dictionary.

Aharon, I., N. Etcoff, et al. (2001). "Beautiful faces have variable reward value: fMRI and behavioral evidence." Neuron 32(3): 537-51.

Angeli, A. D., A. Sutcliffe, et al. (2006). Interaction, usability and aesthetics: what influences users' preferences? Proceedings of the 6th conference on Designing Interactive systems. University Park, PA, USA, ACM: 271-280.

Ashton, J. (2011). "Two Problems with a Neuroaesthetic Theory of Interpretation." Nonsite(2).

Back, I. H. (2010). "Commitment bias: mistaken partner selection or ancient wisdom?" Evolution and Human Behavior 31(1): 22-28.

Bangor, A., P. T. Kortum, et al. (2008). "An Empirical Evaluation of the System Usability Scale." International Journal of Human-Computer Interaction 24(6): 574 - 594.

Bangor, A., P. T. Kortum, et al. (2009). "Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale." Journal of Usability Studies 4(3): 114-123.

Bauerly, M. and Y. Liu (2006). "Computational modeling and experimental investigation of effects of compositional elements on interface and design aesthetics." International Journal of Human-Computer Studies 64(8): 670-682.

Bauerly, M. and Y. Liu (2008). "Effects of Symmetry and Number of Compositional Elements on Interface and Design Aesthetics." International Journal of Human-Computer Interaction 24(3): 275-287.

Baumeister, R. F., E. Bratslavsky, et al. (2001). "Bad is stronger than good." Review of General Psychology Review of General Psychology 5(4): 323-370.

Ben-Bassat, T. M. J. and N. Tractinsky (2006). "Economic and Subjective Measures of the Perceived Value of Aesthetics and Usability." ACM transactions on computer-human interaction : a publication of the Association for Computing Machinery. 13(2): 210.

Bi, L., X. Fan, et al. (2011). "Effects of symmetry and number of compositional elements on Chinese users' aesthetic ratings of interfaces: Experimental and modeling investigations." Int. J. Hum.-Comput. Interact. International Journal of Human-Computer Interaction 27(3): 245-259.

Brady, L. and C. Phillips (2003). "Aesthetics and usability: a look at color and balance." Usability News 5(1): 2003.

Brooke, J. (1996). SUS: A quick and dirty usability scale. Usability evaluation in industry. P. W. Jordan, B. Weerdmeester, A. Thomas and I. L. McLelland, Taylor and Francis.

Cela-Conde, C. J., E. Munar, et al. (2009). "Sex-related similarities and differences in the neural correlates of beauty." Proc. Natl. Acad. Sci. U. S. A. Proceedings of the National Academy of Sciences of the United States of America 106(10): 3847-3852.

Chawda, B., B. Craft, et al. (2005). Do attractive things work better? An exploration of search tool visualizations. Human-Computer Interaction, Edinburgh, Scottland, UK.

Chen, N. Y., D. R. Shaffer, et al. (1997). "On Physical Attractiveness Stereotyping in Taiwan: A Revised Sociocultural Perspective." The Journal of Social Psychology 137(1): 117 - 124.

Chin, J. P., V. A. Diehl, et al. (1988). Development of an instrument measuring user satisfaction of the human-computer interface. Proceedings of the SIGCHI conference on Human factors in computing systems. Washington, D.C., United States, ACM: 213-218.

Coursaris, C. K., S. J. Sweirenga, et al. (2008). "An empirical investigation of color temperature and gender effects on web aesthetics." Journal of Usability Studies 3(3): 103-117.

Cyr, D., M. Head, et al. (2006). "Design aesthetics leading to m-loyalty in mobile commerce." Information & Management 43(8): 950-963.

Cyr, D., M. Head, et al. (2010). "Colour appeal in website design within and across cultures: A multi-method evaluation." International Journal of Human-Computer Studies 68(1-2): 1-21.

Dai, Y., B. Chakraborty, et al. (2011). Kansei engineering and soft computing : theory and practice. Hershey, PA, Engineering Science Reference.

De Angeli, A., A. Sutcliffe, et al. (2006). Interaction, usability and aesthetics: what influences users' preferences?, ACM.

Dion, K., E. Berscheid, et al. (1972). "WHAT IS BEAUTIFUL IS GOOD." Journal of Personality & Social Psychology 24(3): 285-290.

Edell, J. A. and M. C. Burke (1987). "The Power of Feelings in Understanding Advertising Effects." Journal of Consumer Research 14(3): 421-433.

Edelman, S. (2008). Computing the mind : how the mind really works. Oxford; New York, Oxford University Press.

Ellis, L. and C. Ficek (2001). "Color preferences according to gender and sexual orientation." Personality and Individual Differences 31(8): 1375-1379.

Galesic, M. and M. Bosnjak (2009). "Effects of Questionnaire Length on Participation and Indicators of Response Quality in a Web Survey." Public Opinion Quarterly 73(2): 349-360.

Griffin, A. M. and J. H. Langlois (2006). "Stereotype Directionality and Attractiveness Stereotyping: Is Beauty Good or is Ugly Bad?" Social Cognition 24(2): 187-206.

Haig, A. and T. Whitfield (2001). "Predicting the aesthetic performance of Web sites: What attracts people." National School of Design, Swinburne University of Technology, Australia.

Hall, R. H. and P. Hanna (2004). "The impact of web page text-background colour combinations on readability, retention, aesthetics and behavioural intention." Behaviour & Information Technology 23(3): 183-195.

Hamermesh, D. S., J. E. Biddle, et al. (1993). Beauty and the Labor Market. Cambridge, Mass., National Bureau of Economic Research.

Hancock, P., P. Aaron, et al. (2005). "Hedonomics: The Power of Positive and Pleasurable Ergonomics." Ergonomics in Design 13: 8-14.

Hartmann, J., A. Sutcliffe, et al. (2007). Investigating attractiveness in web user interfaces. Proceedings of the SIGCHI conference on Human factors in computing systems. San Jose, California, USA, ACM: 387-396.

Hassenzahl, M. (2004). "The interplay of beauty, goodness, and usability in interactive products." Hum.-Comput. Interact. 19(4): 319-349.

Hassenzahl, M. and A. Monk (2010). "The Inference of Perceived Usability From Beauty." Human-Computer Interaction 25(3): 235-260.

Iaria, G., C. J. Fox, et al. (2008). "The contribution of the fusiform gyrus and superior temporal sulcus in processing facial attractiveness: Neuropsychological and neuroimaging evidence." Neuroscience Neuroscience 155(2): 409-422.

Kampf, M. (2004). What is Usable is Usable Information Science, Utrecht University. Master's Degree: 26.

Kant, I. and W. S. Pluhar (1997). Critique of judgement. Indianapolis, Ind., Hackett.

Karvonen, K. (2000). The beauty of simplicity. Proceedings on the 2000 conference on Universal Usability. Arlington, Virginia, United States, ACM: 85-90.

Kawabata, H. and S. Zeki (2004). "Neural Correlates of Beauty." Journal of Neurophysiology 91(4): 1699-1705.

Kim, J., J. Lee, et al. (2003). "Designing emotionally evocative homepages: an empirical study of the quantitative relations between design factors and emotional dimensions." International Journal of Human-Computer Studies 59(6): 899-940.

Koscinski, K. (2009). "Current status and future directions of research on facial attractiveness." Anthropol. Rev. Anthropological Review 72(1): 45-65.

Kuehni, R. G. (2005). "Color an introduction to practice and principles." from http://public.eblib.com/EBLPublic/PublicView.do?ptiID=221255.

Kurosu, M. and K. Kashimura (1995). Apparent usability vs. inherent usability: experimental analysis on the determinants of the apparent usability. Conference companion on Human factors in computing systems. Denver, Colorado, United States, ACM: 292-293.

Lai, C.-Y., P.-H. Chen, et al. (2010). "Computational models and experimental investigations of effects of balance and symmetry on the aesthetics of text-overlaid images." International Journal of Human-Computer Studies 68(1-2): 41-56.

Lavie, T. and N. Tractinsky (2004). "Assessing dimensions of perceived visual aesthetics of web sites." International Journal of Human-Computer Studies 60(3): 269-298.

Lewis, J. R. (1995). "IBM Computer Usability Satisfaction Questionnaires: Psychometric Evaluation and Instructions for Use." International journal of human-computer interaction. 7(1): 57.

Lindgaard, G., G. Fernandes, et al. (2006). "Attention web designers: You have 50 milliseconds to make a good first impression!" Behaviour & Information Technology 25(2): 115.

Ling, J. and P. van Schaik (2002). "The effect of text and background colour on visual search of Web pages." Displays 23(5): 223-230.

Loyau, A., D. Gomez, et al. (2007). "Iridescent structurally based coloration of eyespots correlates with mating success in the peacock." Behavioral Ecology 18(6): 1123-1131.

Moshagen, M. and M. T. Thielsch (2010). "Facets of visual aesthetics." International Journal of Human-Computer Studies 68(10): 689-709.

Moss, G., R. Gunn, et al. (2006). "Some men like it black, some women like it pink: consumer implications of differences in male and female website design." Journal of Consumer Behaviour Journal of Consumer Behaviour 5(4): 328-341.

Moss, G. A. and R. W. Gunn (2009). "Gender differences in website production and preference aesthetics: Preliminary implications for ICT in education and beyond." Behav Inf Technol Behaviour and Information Technology 28(5): 447-460.

Ngo, D. C. L., L. S. Teo, et al. (2000). A mathematical theory of interface aesthetics.

Ngo, D. C. L., L. S. Teo, et al. (2003). "Modelling interface aesthetics." Information Sciences 152: 25-46.

Nielsen, J. "Usability 101: Introduction to Usability." Retrieved March 24, 2011, from http://www.useit.com/alertbox/20030825.html.

Nielsen, J. (1993). Usability engineering. Boston, Academic Press.

Nielsen, J. (2004). Designing web usability, Pearson Education.

Nielsen, J. and R. L. Mack (1994). Usability inspection methods. New York, Wiley.

Norman, D. A. (2004). Emotional design : why we love (or hate) everyday things. New York, Basic Books.

Pandir, M. and J. Knight (2006). "Homepage aesthetics: The search for preference factors and the challenges of subjectivity." Interacting with Computers 18(6): 1351-1370.

Papachristos, E., N. Tselios, et al. (2006). Modeling Perceived Value of Color in Web Sites. Advances in Artificial Intelligence. G. Antoniou, G. Potamias, C. Spyropoulos and D. Plexousakis, Springer Berlin / Heidelberg. 3955: 567-570.

Pearson, J. M. and A. M. Pearson (2008). "An exploratory study into determining the relative importance of key criteria in web usability: A multi-criteria approach." J. Comput. Inf. Syst. Journal of Computer Information Systems 48(4): 115-128.

Quinn, J. M. and T. Q. Tran (2010). Attractive phones don't have to work better: independent effects of attractiveness, effectiveness, and efficiency on perceived usability. Proceedings of the 28th international conference on Human factors in computing systems. Atlanta, Georgia, USA, ACM: 353-362.

Ramachandran and Hirstein (1999). "The science of art: a neurological theory of aesthetic experience." Journal of Consciousness Studies 6(6): 15-51.

Rau, P.-L. P., Q. Gao, et al. (2007). "The Effect of Rich Web Portal Design and Floating Animations on Visual Search." International Journal of Human-Computer Interaction 22(3): 195-216.

Reimer, J. (2005) "A History of the GUI." Ars Technica.

Rhodes, G. (2006). "The Evolutionary Psychology of Facial Beauty." Annual Review of Psychology 57(1): 199-226.

Rhodes, G. and L. A. Zebrowitz (2002). Facial attractiveness : evolutionary, cognitive, and social perspectives. Westport, Conn., Ablex.

Rozin, P. and E. B. Royzman (2001). "Negativity Bias, Negativity Dominance, and Contagion." Personality and Social Psychology Review 5(4): 296-320.

Salkind, L. and N. J. Salkind (1997). "Gender and Age Differences in Preference for Works of Art." Studies in Art Education 38(4): 246-56.

Sauro, J. (2010, January 18). "Can you use the SUS for websites?" Retrieved March 24, 2011, from http://www.measuringusability.com/blog/sus-websites.php.

Sauro, J. (2011). "The Standardized Universal Percentile Rank Questionnaire (SUPR-Q)." Retrieved March 24, 2011, from http://www.suprq.com/index.php.

Schenkman, B. N. and F. U. Jönsson (2000). "Aesthetics and preferences of web pages." Behaviour & Information Technology 19(5): 367-377.

Schmidt, K. E., Y. Liu, et al. (2009). "Webpage aesthetics, performance and usability: Design variables and their effects." Ergonomics 52(6): 631-643.

Schrepp, M., T. Held, et al. (2006). "The influence of hedonic quality on the attractiveness of user interfaces of business management software." Interacting with Computers 18(5): 1055-1069.

Schütte, S. (2005). Engineering emotional values in product design : Kansei engineering in development. Linköping, Dept. of Mechanical Engineering, Univ.

Shieh, K.-K. and C.-C. Lin (2000). "Effects of screen type, ambient illumination, and color combination on VDT visual performance and subjective preference." International Journal of Industrial Ergonomics 26(5): 527-536.

Simon, S. J. (2000). "The impact of culture and gender on web sites: an empirical study." SIGMIS Database 32(1): 18-37.

Sonderegger, A. and J. Sauer (2010). "The influence of design aesthetics in usability testing: Effects on user performance and perceived usability." Applied Ergonomics 41(3): 403-410.

Sutcliffe, A. and A. De Angeli (2005). "Assessing Interaction Styles in Web User Interfaces." Lecture notes in computer science.(3585): 405-417.

Tarasewich, P., H. Z. Daniel, et al. (2001). "Aesthetics and web site design." Quarterly Journal of Electronic Commerce 2: 67-82.

Thielsch, M. T. and G. Hirschfeld (2010). "High and low spatial frequencies in website evaluations." Ergonomics 53(8): 972.

Tractinsky, N. (1997). Aesthetics and apparent usability: empirically assessing cultural and methodological issues. CHI '97: Proceedings of the SIGCHI conference on Human factors in computing systems, Atlanta, Georgia, United States, ACM.

Tractinsky, N. (2000). "What is beautiful is usable." Interacting with Computers 13(2): 127-145.

Tuch, A. N., J. A. Bargas-Avila, et al. (2009). "Visual complexity of websites: Effects on users' experience, physiology, performance, and memory." International Journal of Human-Computer Studies 67(9): 703-715.

Tullis, T. S. S., J. N. (2004). A Comparison of Questionnaires for Assessing Website Usability. Usability Professionals Association (UPA). Minneapolis, USA.

van Schaik, P. and J. Ling (2008). "Modelling user experience with web sites: Usability, hedonic value, beauty and goodness." Special Issue: On the Abuse and Misuse of Social Agents 20(3): 419-432.

van Schaik, P. and J. Ling (2009). "The role of context in perceptions of the aesthetics of web pages over time." International Journal of Human-Computer Studies 67(1): 79-89.

Wauters, R. (July 2010) "Digg Vs. Reddit: The Infographic." TechCrunch.

Winston, J. S., J. O'Doherty, et al. (2007). "Brain systems for assessing facial attractiveness." Neuropsychologia 45(1): 195-206.

Zeki, S. (2001). "Essays on science and society. Artistic creativity and the brain." Science (New York, N.Y.) 293(5527): 51-2.

Zheng, X. S., I. Chakraborty, et al. (2009). Correlating low-level image statistics with users - rapid aesthetic and affective judgments of web pages. Proceedings of the 27th international conference on Human factors in computing systems. Boston, MA, USA, ACM: 1-10.